

**UNIVERSIDADE FEDERAL DO PAMPA**

**BRUNO CAMPOS DE VASCONCELLOS**

**MÉTODO APLICADO AO  
MONITORAMENTO REMOTO DE  
ANIMAIS BASEADO EM  
AEROLEVANTAMENTO COM VANT E  
APRENDIZAGEM PROFUNDA**

**Bagé  
2019**

**BRUNO CAMPOS DE VASCONCELLOS**

**MÉTODO APLICADO AO  
MONITORAMENTO REMOTO DE  
ANIMAIS BASEADO EM  
AEROLEVANTAMENTO COM VANT E  
APRENDIZAGEM PROFUNDA**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Orientador: Leonardo Bidese de Pinho  
Coorientador: Leandro Bochi da Silva Volk

**Bagé  
2019**

V331m Vasconcellos, Bruno Campos de

Método Aplicado ao Monitoramento Remoto de Animais Baseado em Aerolevanteamento com VANT e Aprendizagem Profunda / Bruno Campos de Vasconcellos.

71 p.

Dissertação (Mestrado) – Universidade Federal do Pampa, Campus Bagé, Programa de Pós-Graduação em Computação Aplicada, 2019.

“Orientação: Leonardo Bidese de Pinho”.

1. Pecuária de precisão. 2. Sensoriamento remoto. 3. Sistemas de informação geográfica. 4. Redes neurais convolucionais. 5. Bancos de dados espaço-temporais.  
I. Título.

**BRUNO CAMPOS DE VASCONCELLOS**

**MÉTODO APLICADO AO MONITORA-  
MENTO REMOTO DE ANIMAIS BASE-  
ADO EM AEROLEVANTAMENTO COM  
VANT E APRENDIZAGEM PROFUNDA**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de concentração: Tecnologias para a produção agropecuária


Dissertação defendida e aprovada em: 1 de Março de 2019.

Banca examinadora:



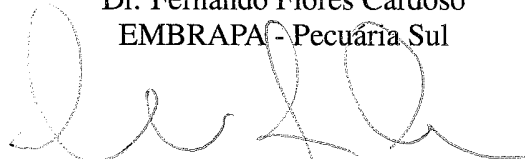
---

Prof. Dr. Leonardo Bidese de Pinho  
Orientador




---

Dr. Fernando Flores Cardoso  
EMBRAPA - Pecuária Sul



---

Prof. Dr. Milton Roberto Heinen  
UNIPAMPA - Campus Bagé



---

Prof. Dr. Paulo Roberto Ferreira Júnior  
UFPEL - CDTEC

## **AGRADECIMENTO**

Agradeço a Deus por me fornecer energia para chegar até aqui de alguma maneira. Agradeço ao meu filho e esposa por suportarem minha ausência para que eu pudesse dar continuidade a esse trabalho. Agradeço a CCANet que acreditou em mim nessa oportunidade de cursar um mestrado. Agradeço ao orientador, Professor Leonardo Pinho, que foi incansável na orientação em madrugadas intermináveis. Agradeço ao coorientador, Professor Leandro Volk, pelas dicas e paciência nos aerolevamentos nas áreas experimentais. Agradeço aos Pesquisadores Naylor Perez e José Pedro Trindade pelo apoio nos aerolevamentos, estendido ao graduando Anderson Soares.

## RESUMO

Uma das demandas necessárias na pecuária de manejo extensivo é a de contagem de animais em áreas de dezenas de hectares, onerosa quando realizada de forma manual e local. Neste contexto, este trabalho propõe e discute a eficácia de um método semiautônomo, não invasivo, para identificação remota de animais a campo, aplicável a sistemas de pecuária de precisão. O método foi concebido a partir de uma metodologia de pesquisa aplicada experimental, exploratória no seu início e posteriormente explicativa, tendo como base técnicas de sensoriamento remoto que incluem processos de coleta de imagens por aerolevantamento com câmera RGB embarcada em veículo aéreo não tripulado, persistência das imagens obtidas por meio de armazenamento em bancos de dados espaço-temporal e processamento das imagens armazenadas para a construção de um ortomosaico da propriedade rural sucedida pela aplicação de processos de descoberta de padrões, fazendo uso de aprendizagem de máquina, em especial de redes neurais convolucionais. Conforme experimentos realizados em um cenário real, o método demonstrou ser eficaz, sendo capaz de, a partir de diferentes coletas de imagens feitas com um VANT a 100 m, detectar e contar animais com acurácia de até 92%, sendo ainda possível identificar a posição geográfica aproximada dos animais a campo.

**Palavras-chave:** Pecuária de precisão. Sensoriamento remoto. Sistemas de informação geográfica. Redes neurais convolucionais. Bancos de dados espaço-temporais.

## ABSTRACT

One of the main demands in the extensive livestock management is the counting of animals in areas of tens of hectares, costly when performed manually and locally. In such context, this work proposes and discusses the effectiveness of a semi-autonomous, non-invasive method for remote identification of animals in the field, applicable to precision livestock farming systems. The method was conceived from an applied experimental methodology, exploratory at first, explanatory later, based on remote sensing techniques that include image collection processes by aerial surveying with a RGB camera embedded in an unmanned aerial vehicle, the persistence of the images obtained by means of storing spatiotemporal data, and processing of stored images to generate an orthomosaic map covering the rural property succeeded by the application of pattern discovery processes, making use of machine learning, especially convolutional neural networks. According to experiments carried out in a real scenario, the method proved to be effective, being able to detect and count animals with an accuracy of up to 92% from different sets of images collected using UAV at 100 m, and it is still possible to identify the position the animals to the field.

**Keywords:** Precision Livestock. Remote Sensing. Geographic Information Systems. Convolutional Neural Networks. Space-time Databases.

## LISTA DE FIGURAS

Figura 1	Diversas resoluções de imagens e suas perdas .....	18
Figura 2	Esquema de funcionamento do algoritmo FAST R-CNN. ....	25
Figura 3	Esquema de funcionamento do algoritmo SSD .....	26
Figura 4	RNA multilayer perceptron .....	29
Figura 5	Imagem antes da aplicação do <i>kernel</i> 3x3 .....	33
Figura 6	Imagem após aplicação do <i>kernel</i> 3x3.....	33
Figura 7	Subamostragem .....	34
Figura 8	Arquitetura de uma CNN.....	35
Figura 9	Passo a passo do método proposto.....	40
Figura 10	Linha de voo e GSD com variação de altitude .....	44
Figura 11	Linhas de voo e sobreposição .....	45
Figura 12	Ortomosaico Voo 1 .....	48
Figura 13	Vacas e geração de sombra .....	49
Figura 14	Ortomosaico Voo 2 .....	50
Figura 15	Modelo de dados HBase .....	50
Figura 16	Passo a passo do método proposto.....	52
Figura 17	Imagem com anotações em animais .....	54
Figura 18	Imagem com detecção e localização de animais referente ao Experimento 1.....	58
Figura 19	Imagem com detecção e localização de animais referente ao Experimento 2.....	59
Figura 20	Imagem com 18 animais a 10 m de altitude referente ao Experimento 4.....	60
Figura 21	Relação entre acurácia e épocas .....	61
Figura 22	Relação entre a melhora da acurácia e IoU .....	62
Figura 23	Imagem com localização geográfica aproximada dos animais.....	63
Figura 24	Relação entre acurácia e tempo nos ambientes 1 (GPU) e 2 (CPU).....	64



## LISTA DE TABELAS

Tabela 1	Experimento 1 incluindo imagens do primeiro voo com 1000 épocas.....	57
Tabela 2	Experimento 2 incluindo imagens do segundo voo com 1000 épocas .....	58
Tabela 3	Experimento 3 incluindo imagens dos dois voos com 1000 épocas.....	59

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>10</b>
<b>1.1 Objetivos e questões de pesquisa</b> .....	<b>12</b>
<b>1.2 Metodologia de pesquisa</b> .....	<b>12</b>
<b>1.3 Organização do texto</b> .....	<b>13</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>14</b>
<b>2.1 Sensoriamento remoto</b> .....	<b>14</b>
<b>2.2 Coleta de imagens aéreas</b> .....	<b>17</b>
2.2.1 Aerofotogrametria.....	17
2.2.2 Ortomosaico .....	19
2.2.3 Plano de voo.....	20
<b>2.3 Armazenamento de imagens aéreas</b> .....	<b>20</b>
2.3.1 Banco de dados espaço-temporais .....	20
<b>2.4 Processamento de imagens aéreas</b> .....	<b>21</b>
2.4.1 Conceitos básicos.....	22
2.4.2 Visão computacional.....	23
2.4.2.1 Algoritmos de localização de objetos em imagens .....	24
2.4.3 Aprendizado de máquina .....	27
2.4.3.1 Redes neurais artificiais .....	27
2.4.3.2 Aprendizado profundo .....	30
<b>3 TRABALHOS RELACIONADOS</b> .....	<b>37</b>
<b>3.1 Reconhecimento de animais em imagens</b> .....	<b>37</b>
<b>3.2 Armazenamento espaço-temporal de imagens</b> .....	<b>38</b>
<b>4 MÉTODO PROPOSTO</b> .....	<b>40</b>
<b>4.1 Primeira Etapa</b> .....	<b>40</b>
<b>4.2 Segunda Etapa</b> .....	<b>41</b>
<b>4.3 Terceira Etapa</b> .....	<b>41</b>
<b>4.4 Quarta Etapa</b> .....	<b>41</b>
<b>5 METODOLOGIA EXPERIMENTAL</b> .....	<b>43</b>
<b>5.1 Materiais e métodos</b> .....	<b>43</b>
5.1.1 Área mapeada.....	43
5.1.2 Equipamentos utilizados .....	43
5.1.3 Imageamento da área .....	44
5.1.4 Transmissão de dados .....	46
5.1.5 Ambiente computacional .....	46
5.1.6 Processamento de imagens .....	47
5.1.7 Armazenamento de imagens .....	47
<b>5.2 Experimentos</b> .....	<b>48</b>
5.2.1 Primeiro experimento.....	53
5.2.2 Segundo experimento.....	53
5.2.3 Terceiro experimento .....	54
5.2.4 Quarto experimento .....	54
5.2.5 Desempenho computacional dos experimentos .....	55
<b>5.3 Métricas</b> .....	<b>55</b>
<b>6 RESULTADOS E DISCUSSÕES</b> .....	<b>57</b>
<b>7 CONCLUSÕES</b> .....	<b>65</b>
<b>REFERÊNCIAS</b> .....	<b>68</b>

## 1 INTRODUÇÃO

De acordo com a FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA (2017), a Pecuária no Rio Grande do Sul é detentora do sexto maior rebanho de bovinos, do segundo maior rebanho de equinos e do maior rebanho de ovinos do território nacional. Em 2014, o valor bruto da produção pecuária do RS totalizou R\$ 15,8 bilhões. Além da bovinocultura de corte, contribuem efetivamente para a composição desse valor a atividade leiteira, a avicultura e a suinocultura.

Nas últimas décadas, o desenvolvimento e a modernização da tecnologia da Pecuária brasileira têm tornado o segmento competitivo, exigindo habilidades de gestão, profissionalismo e especialização. A automação e outras novas tecnologias estão, cada vez mais, presentes no dia a dia rural, com o objetivo de aumentar a capacidade e o sucesso da produção (TEIXEIRA; HESPANHOL, 2014).

Neste sentido, a pecuária de precisão, termo definido pela EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (2018) como sendo “a gestão da produção que, a partir do uso de conhecimentos e tecnologias variadas, busca entender e intervir na variação que existe, tanto no ambiente de produção, como no rebanho”, busca melhorar processos anteriormente considerados como homogêneos, levando em consideração os diversos estados existentes na produção de animais. Entretanto, em ambientes de manejo extensivo, a observação humana das diferentes variáveis que compõem este complexo sistema requer tempo significativo, dependendo das características da propriedade (principalmente a extensão) e dos animais (em especial a quantidade/densidade). Outra demanda que envolve fatores econômicos a pecuaristas é a necessidade de controlar o estoque de animais. Esta representa uma das atividades mais importantes, implicando a necessidade de realizar um inventário de animais com frequência diária ou semanal. As variações no estoque de animais podem ocorrer por morte, nascimento, abigeato (furto ou roubo de animais) e transferência entre produtores por meio de venda (BARBALHO *et al.*, 2005).

Entre as tecnologias disponíveis com potencial para auxiliar os pecuaristas no manejo extensivo se destacam os Veículos Aéreos Não Tripulados (VANT), também conhecidos como Aeronaves Remotamente Pilotadas (ARP) ou mais simplesmente como *drones*, os quais são aeronaves capazes de sobrevoar áreas de interesse, transportando diversos tipos de sensores, de forma semiautônoma. No que se refere à pecuária de precisão, Poloni (2016) afirma que, entre as funcionalidades mais comuns de *drones*, estão o acompanhamento da pastagem, o manejo do gado, a telemetria, a contagem e observação dos cochos,

a busca por animais perdidos, a contagem do rebanho e a quantidade de animais que estão no pasto.

Dentro do contexto particular das aplicações de contagem e localização do rebanho na pastagem, Mulero-Pázmány *et al.* (2015) realizaram um estudo para avaliar a utilidade dos *drones* em obter imagens capturadas por câmera fotográfica digital a fim de modelar a distribuição do rebanho bovino comparada aos resultados obtidos usando rastreadores fixos nos animais (tipo colar), concluindo que o seu uso consiste em uma prática não-invasiva, capaz de fornecer dados espaciais precisos, úteis para pesquisa ecológica, manejo animal e planejamento de pastagens. Por outro lado, a revisão da literatura realizada mostrou que, de forma geral, a utilização de sensoriamento via visão de máquina ou processamento de imagem é uma solução ainda recente no contexto da pecuária de precisão, mas pertinente, uma vez que permite a coleta de dados sobre grandes grupos de animais sem necessidade de uma estrutura de mesma proporção, deixando o custo ligado proporcionalmente ao processamento de dados e não à coleta de informações.

Como exemplo, Moroz e JR (2016) desenvolveram uma solução, baseada em circuitos fechados de gravação de imagens (CFTV), capaz de monitorar a distribuição dos animais no ambiente e, aliada a câmeras específicas para detecção de movimento (Sensor Kinect, desenvolvido pela Microsoft), efetuar a caracterização de comportamento anormal do rebanho. A contagem de animais por meio de VANT também está relacionada à detecção de comportamento animal, a qual envolve toda interação entre o ser e o ambiente em si. Neste contexto, existe um amplo espectro de pesquisas e aplicações desenvolvidas, tanto em nível de dispositivos, quanto algoritmos, buscando quantificar, de maneira indireta, o bem-estar e a saúde do animal (MOROZ; JR, 2016).

Complementarmente, é importante ressaltar que a resolução espacial dos sensores remotos orbitais (satélites) dificulta a identificação de animais em razão de suas limitações de resolução. Por outro lado, os VANT podem produzir imagens com resolução de 20 cm ou menos, onde os animais são reconhecíveis e podem ser contados (REY, 2016). Além disso, os VANT, ao contrário dos satélites, não apresentam uma limitação significativa no que se refere à resolução temporal, associada à frequência de revisitação da área de interesse. Para ilustrar a limitação do uso de satélites para a contagem de animais, cabe destacar, como exemplo, a rede de satélites do projeto Sentinel-2, considerada estado-da-arte, que apresenta resolução de 10 metros por pixel e capacidade de revisita de cinco dias. Portanto, características insuficientes para a construção de uma solução eficaz para o problema exposto.

## 1.1 Objetivos e questões de pesquisa

Partindo das justificativas expostas, o presente trabalho tem como objetivo principal desenvolver um método capaz de monitorar remotamente animais e realizar de forma eficaz a contagem autônoma do rebanho a campo, a partir de imagens (*Red Green Blue*) RGB coletadas por meio de VANT, sendo posteriormente transmitidas, armazenadas e processadas com a aplicação de técnicas de visão computacional.

Como consequência, almeja-se uma forma não invasiva para obtenção da contagem e posicionamento dos animais no campo com uma acurácia e desempenho suficientes para diversas aplicações, em especial para estudos que permitam uma melhor compreensão sobre o comportamento do rebanho.

Neste contexto, como parte do processo investigativo delineado, tenta-se responder quatro principais questões de pesquisa:

- A qualidade das imagens RGB coletadas por meio de um VANT é suficiente para identificar os animais a campo?
- É possível mensurar o impacto de fatores que influenciam a coleta de imagens por meio de VANT, tais como a altitude (maior ou menor resolução espacial), o terreno (plano, ondulado ou acidentado), a vegetação (presença de árvore e arbustos) e a luminosidade (horário e nebulosidade), na eficácia do método de identificação dos animais?
- A visão computacional consegue suprir a necessidade da contagem de animais sob a perspectiva de processamento computacional?
- Há técnicas capazes de realizar essa tarefa com acurácia maior que 90%, dentro do contexto de visão computacional?

## 1.2 Metodologia de pesquisa

O presente trabalho se ampara em uma metodologia de pesquisa de natureza aplicada, capaz de “gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos” (PRODANOV; FREITAS, 2013). No que se refere aos seus objetivos, adotou inicialmente uma abordagem exploratória, visando ampliar o domínio do problema, soluções existentes e técnicas aplicáveis para a concepção de uma proposta de solução amparada por um referencial teórico relevante. Posteriormente, adotou uma

abordagem explicativa, objetivando “registrar, analisar, classificar e interpretar” o efeito da aplicação do método proposto, sobretudo buscando identificar os principais fatores que afetam o seu desempenho.

Além disso, o trabalho poder ser classificado como experimental conforme os seus procedimentos técnicos, já que envolve a seleção de variáveis capazes de influenciar a eficiência do método, definição de formas de controle e de observação do efeito das variáveis nos resultados obtidos por meio de métricas quantitativas em um cenário realista de aplicação. O resultado desta metodologia permite reunir um conjunto de dados experimentais que permite “demonstrar o modo ou as causas pelas quais um fato é produzido, proporcionando, assim, o estudo de suas causas e seus efeitos”.

### **1.3 Organização do texto**

O restante do texto está estruturado como segue. No Capítulo 2 são apresentados os fundamentos teóricos do trabalho, divididos conforme as etapas do processo necessário para a identificação de animais em imagens decorrentes de aerolevantamentos com VANT. No Capítulo 3 são sintetizados os trabalhos correlatos identificados a partir da revisão da literatura realizada com foco no tema do trabalho. No Capítulo 4 é apresentado em detalhes o método proposto. No Capítulo 5 são descritos os materiais e métodos utilizados na metodologia experimental adotada para avaliação empírica do método proposto em um cenário real. No Capítulo 6 são apresentados e discutidos os resultados experimentais obtidos com diferentes parametrizações do método. No Capítulo 7 são apresentadas as conclusões acerca do potencial do método e das questões de pesquisa, a luz dos resultados experimentais obtidos, bem como são elencados possíveis trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados os fundamentos teóricos do trabalho, iniciando-se pela apresentação dos referenciais sobre aspectos gerais de Sensoriamento Remoto. Na sequência, são apresentados elementos específicos sobre as etapas de coleta, armazenamento e processamento, os quais servem de base para o desenvolvimento do método proposto neste trabalho.

### 2.1 Sensoriamento remoto

Meneses e Almeida (2012) definiram sensoriamento remoto como “uma técnica de obtenção de imagens dos objetos da superfície terrestre sem que haja um contato físico de qualquer espécie entre o sensor e o objeto”. Alguns preceitos sobre os quais o sensoriamento remoto é baseado são:

- Exigência: ausência de matéria no espaço entre o objeto e o sensor;
- Consequência: a informação do objeto é possível de ser transportada pelo espaço vazio;
- Processo: o elo de comunicação entre o objeto e o sensor é a radiação eletromagnética, a única forma de energia capaz de se transportar pelo espaço.

Neste escopo, o sensoriamento remoto é explícito em afirmar que o objeto imageado é registrado pelo sensor por meio de medições da radiação eletromagnética, tal como a luz solar refletida da superfície de qualquer objeto. Nenhum outro tipo de sensor que obtenha imagens que não seja pela detecção da radiação eletromagnética deve ser classificado como sensoriamento remoto (MENESES; ALMEIDA, 2012).

A radiação eletromagnética (REM), onda e energia, propaga-se pelo espaço vazio, tanto a partir da Terra quanto do Sol. Com base nessa afirmação entende-se que, para um imageamento por sensoriamento remoto, seja imprescindível a incidência da REM nos objetos e, ainda, que cada objeto responda de uma maneira diferente, resultando na chamada assinatura espectral dos objetos.

Meneses e Almeida (2012) afirmam que, em média, “47% do fluxo de radiação que incide na superfície terrestre é absorvido pelos materiais da própria superfície, 37% é refletido pela Terra (incluindo nuvens e atmosfera) e 17% é absorvido pela atmosfera”. A parte da REM que interage diretamente com a atmosfera sofre dois efeitos, absorção e

espalhamento da radiação, e esse comportamento da atmosfera é questão crucial para o sensoriamento remoto de alta altitude ou orbital. Como consequência, há algumas áreas não cobertas pelo sensoriamento remoto por questões físico-químicas.

Áreas nubladas, por exemplo, são impossíveis de serem imageadas por meio da luz solar, como consequência reduz a área útil da imagem, causado pela nuvem e pela projeção de sua sombra no terreno, existe a interferência na análise quantitativa da imagem. Nesse caso, pixels interferidos por nuvem e sombra são tratados como pixels válidos e contaminam os valores calculados (SILVA; LIPORACE, 2016). Para tanto, torna-se necessário o procedimento de eliminação das nuvens no pós-processamento, conforme destacam Silva e Liporace (2016), existindo metodologias para resolver este problema.

Polidorio *et al.* (2005) apresentam uma técnica para segmentação de corpos d'água, nuvens e sombras em imagens multiespectrais adquiridas pelos satélites CBERS (satélite lançado em parceria entre China e Brasil em outubro de 1999) e LANDSAT (satélite lançado ao espaço em 1972).

Silva, Salviano e Andrade (2007) trazem outra metodologia, voltada para imagens do sensor *Charge-Coupled Device* (CCD) - semicondutor utilizado para captação de imagens formado por um circuito integrado composto por uma matriz de capacitores acoplados cuja capacidade de resolução depende do tamanho e número das suas células fotoelétricas - a bordo do satélite CBERS, onde a correlação nuvem/sombra é indicada para a detecção de nuvens de baixa altitude, tais como nuvens do tipo *Cumulus* e *Stratocumulus* que apresentam sombras possíveis de detectar e, normalmente, ocupam pequenas áreas.

Outra característica das imagens coletadas por sensoriamento remoto é, para que o movimento dos satélites cubram praticamente todas as regiões do Globo, aqueles acompanham a Terra no movimento de translação, mas não no de rotação. As passagens em dias sucessivos não são coincidentes, assim, o satélite passa a imagear outras faixas e só voltam a visitar uma mesma área após um certo período de tempo. O satélite LANDSAT demora 16 dias para voltar a uma mesma faixa, o SPOT (Satélite Francês para Observação da Terra, lançado nos anos 80 – melhorou a resolução espacial ) demora 26 dias, o (*National Oceanic and Atmospheric Administration*) NOAA cobre uma mesma faixa quase todos os dias, devido a sua larga faixa de imageamento.

Com períodos orbitais de, aproximadamente 100 a 103 minutos, no caso do LANDSAT, do SPOT e do NOAA, os satélites realizam quatorze voltas inteiras mais uma fração de volta, em torno da Terra, em um período de 24 horas. Handcock *et al.*



(2009) propuseram uma forma de monitoramento do comportamento animal baseada em *Global Position System* (GPS), em conjunto com imagens de sensoriamento remoto baseada em satélites, e ainda uma rede de sensores sem fios. Neste trabalho destaca-se que a integração de dados de telemetria, em particular dados de GPS que registram locais de animais, com dados de sensoriamento remoto, permite que as preferências dos animais sejam diretamente ligadas às medições espacialmente extensas da paisagem a partir de imagens aéreas ou de satélite. Uma vez modelada a relação entre o comportamento animal e a análise espectral e espacial das imagens de sensoriamento remoto, essas relações podem ser estendidas a novas imagens para prever e mapear o provável comportamento animal.

Essas informações podem ser usadas para gerenciar melhor os movimentos dos animais e proteger áreas sensíveis contra a degradação e o excesso de pastagem. Em particular, nota-se que as questões de frequência de coleta das imagens podem ser superadas através da obtenção de dados mais frequentes, por meio de uso de VANT para capturar dados de sensoriamento remoto. Além disso, identifica-se que o trabalho citado precisou de diversas formas para detecção quando o propósito foi monitorar o comportamento animal, deixando evidente que o imageamento baseado no sensoriamento remoto por satélite pode não ser uma opção para o proposto neste trabalho, já que a aquisição das imagens tem limitações temporais, como foi mostrado nos satélites SPOT, LANDSAT e NOAA. Outra evidência é o uso de pós-processamento em relação a interferências atmosféricas como as nuvens, o que não seria viável para efetivar uma automatização por necessitar de muita intervenção humana.

Outras vantagens do VANT para este trabalho, estão relacionadas ao relativo baixo custo de aquisição e fácil manuseio do equipamento, à possibilidade de obtenção de dados de alta resolução espacial Nepomuceno (2016).

Além disso, um aspecto destacado por Rey (2016) é o custo da aquisição de imagens, que dependem da resolução espacial desejada: as imagens de satélite disponíveis gratuitamente têm uma resolução de 30 m, enquanto as imagens de satélites comerciais têm uma resolução de até 20 cm. A legislação de muitos países não permite uma resolução maior, de modo que a limitação é mais legal do que técnica.

## 2.2 Coleta de imagens aéreas

No presente trabalho é exigida uma etapa onde as imagens serão coletadas por meio de um sensor CCD embarcado em um VANT. Em virtude disso, há necessidade de apresentar conceitos e técnicas relevantes aplicadas a este contexto, com destaque para uma das modalidades de aerolevantamentos denominada aerofotogrametria, bem como para a criação de ortomosaico das imagens e ao planejamento do voo.

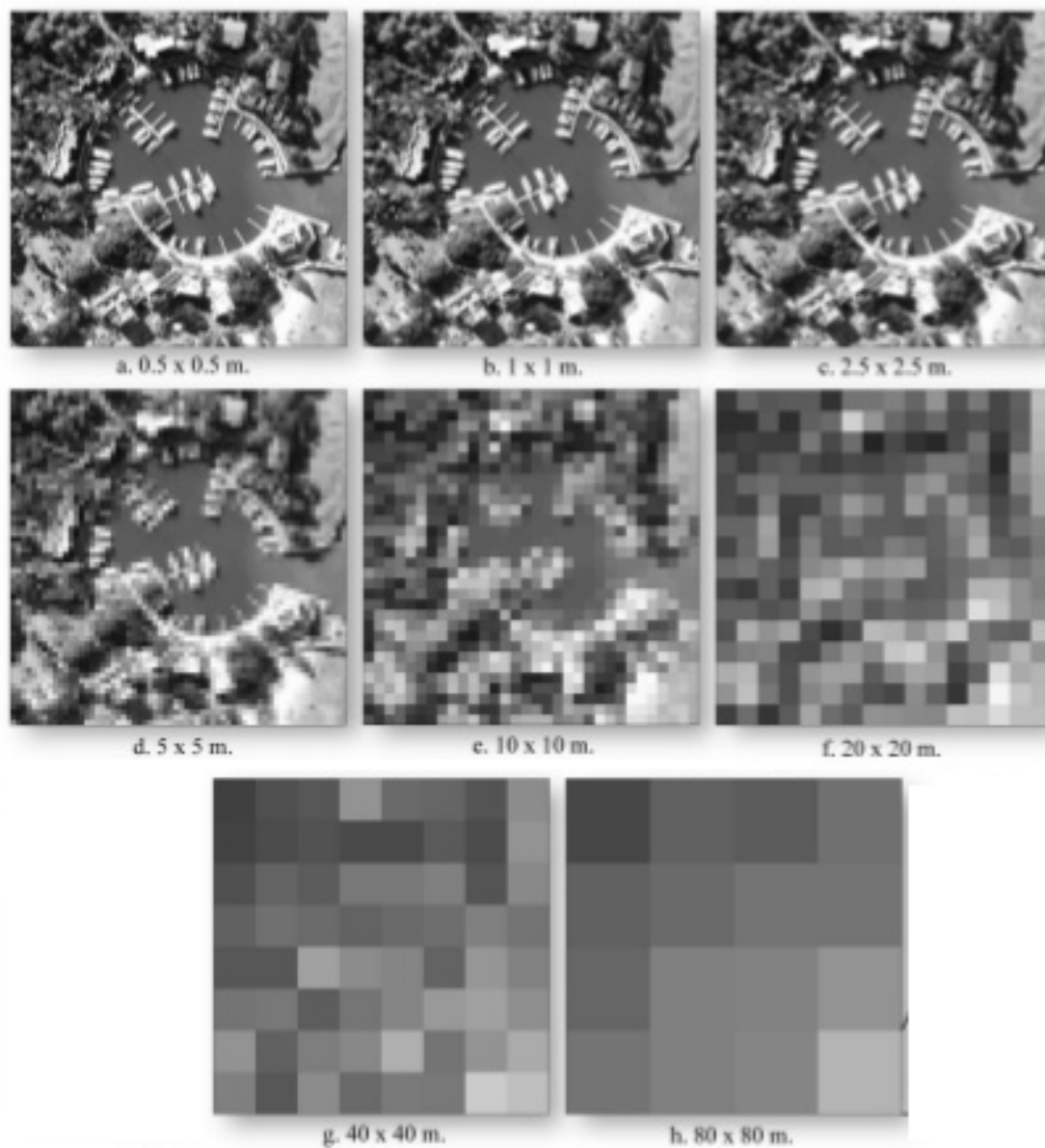
### 2.2.1 Aerofotogrametria

A aerofotogrametria é a ciência e a tecnologia de obter informações seguras acerca de objetos físicos e do meio, através de processos de registro, medição e interpretação das imagens fotográficas, obtendo distâncias aproximadas, áreas e elevações usando fotografias (JENSEN; EPIPHANIO, 2011).

Como resultado da aerofotogrametria, as ortofotos são fotografias que mostram imagens de objetos em suas posições ortográficas verdadeiras. As ortofotos são geometricamente equivalentes a mapas convencionais planimétricos de linhas e símbolos, os quais também mostram as posições ortográficas verdadeiras dos objetos (WOLF, 1983). O processo de criação de um ortofoto depende da existência de um ( modelo digital de elevação) DEM confiável para a área que está sendo mapeada. O DEM fornece dados de altitude. Além disso, estes dados são planimetricamente rigorosos, e assim, os elementos bidimensionais podem ser digitalizados diretamente. Outro modelo para geração de ortofotos, consiste em pontos que modelam o terreno considerando somente os objetos acima do solo como árvores e animais (WOLF, 1983). A aerofotogrametria também exige a definição da escala de uma fotografia vertical e estimar as distâncias terrestres horizontais, visto que as variações entre o resultado de um mapeamento de um objeto pode trazer um resultado inesperado, conforme se pode ver na Figura 1 uma imagem aérea de uma cidade com resoluções entre 0,5 x 0,5 m a 80 x 80 m. Não há dificuldades em interpretabilidade dos dados de 0,5 x 0,5, 1 x 1 m ou 2 x 2 m. Entretanto, a partir da imagem 10 x 10 m esta fica inutilizada para identificação dos elementos urbanos (JENSEN; EPIPHANIO, 2011).

Para obter a definição correta na aerofotogrametria existe a variável *Ground Sample Distance* (GSD), correspondente à “Distância de amostra do solo”. O GSD é a representação do pixel da imagem em unidades de terreno (geralmente em cm). Esta é uma das variáveis mais importantes, sendo a primeira que deve ser definida, pois ela determina a

Figura 1 – Diversas resoluções de imagens e suas perdas



Fonte: Jensen e Epiphonio (2011)

resolução espacial da coleta de dados, ou seja, o seu nível de detalhamento. A escolha do GSD influencia diretamente na qualidade da imagem envolvida na coleta dos dados e na capacidade de mapeamento pois, para aumentar o nível de detalhamento do pixel, deve-se sobrevoar a uma altitude menor, cobrindo uma área menor.

No contexto de VANT, o uso de uma plataforma aérea de levantamento fotogramétrica pode ser operada remotamente, de forma semiautônoma ou totalmente autônoma. Essa plataforma é equipada com um sistema de câmera fotográfica digital comum ou infravermelho, um sistema *Global Navigation Satellite System* (GNSS) para fornecer a posição do centro de cada fotografia, um *Inertial Measurement Unit* (IMU) composto de giroscópios, acelerômetros, altímetros e bússola, que permitem a determinação da orientação exterior aproximada para cada imagem tomada durante o voo, link de rádio e uma pequena CPU que controla tudo. O controle humano é feito por um sistema remoto com o link de rádio (SILVA *et al.*, 2014).

Como são aeronaves capazes de provocar acidentes em pessoas se conduzidas por operadores sem treinamento prévio, os VANT tem regulamentação por meio da Agência Nacional de Aviação Civil (ANAC). Voos até 120 em aeronaves de até 25 kg, são classificados como voos permitidos sem necessidade de registros no órgão (AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL, 2017).

### **2.2.2 Ortomosaico**

Ortomosaico é um mosaico construído a partir de ortofotos sequenciais, com auxílio de pontos de controle obtidos por triangulação. O mosaico gerado com a aquisição de pontos de controle é nomeado de mosaico controlado. Um dos principais algoritmos utilizados para junção das imagens é o *Scale Invariant Feature Transform* (SIFT), capaz de reconhecer áreas correlatas das imagens formando o ortomosaico, obtendo uma imagem com informações da área a ser mapeada. A montagem do ortomosaico entrega um resultado mais refinado, já que a visão da área mapeada é total, ainda que degrade a performance computacional (SILVA *et al.*, 2015).

### 2.2.3 Plano de voo

Além dos equipamentos são necessários vários softwares para o planejamento da missão e para controle durante voo. Os softwares de planejamento de voo fotogramétrico são similares aos da fotogrametria convencional: conhecida a distância focal da câmera, a escala ou GSD, definidas as sobreposições (maiores que na fotogrametria convencional: 70-80% longitudinal e lateral) e as coordenadas dos cantos da área, o software lança em mapa as linhas de voo e estima o tempo necessário para sua conclusão.

Os softwares de controle durante voo monitoram o funcionamento da câmera, velocidade, quantidade de energia (bateria ou combustível) e as regras de segurança para fazer o veículo retornar para o ponto de partida se algo acontecer fora do previsto. Os dados do voo em seguida são introduzidos em softwares de pré-processamento que fazem a ligação entre os instantes de tomada das imagens, GNSS e IMU, gerando o chamado arquivo de log GPS, fundamental para a etapa de processamento fotogramétrico (SILVA *et al.*, 2014).

## 2.3 Armazenamento de imagens aéreas

Esta seção trata do armazenamento das imagens em um banco de dados, necessária em função do grande volume de imagens obtidas a partir do aerolevanteamento por VANT, frequentemente da ordem de Gigabytes. Ainda que seja realizado algum pré-processamento na imagem, armazenar a imagem e dados sobre elas, como número de objetos encontrados na imagem, mostra-se relevante para posteriormente os objetos serem identificados mais rapidamente, bem como mantendo a temporalidade dos dados através de novas coletas de imagens sobre uma determinada área, formando uma série histórica de dados espaço-temporais.

### 2.3.1 Banco de dados espaço-temporais

No contexto deste trabalho, os dados armazenados são imagens *raster* e seus metadados posição global, bandas e tamanho. Na escala temporal espera-se um aumento mensal destes dados, pretende-se admitir o uso de um banco de dados distribuídos capaz de suportar grandes volumes de dados.

Os bancos de dados NoSQL, conforme explicam Ferreira e Oliveira (2014), são capazes de armazenar grandes volumes de dados com alto desempenho. Com o intuito de conseguir aliar alto desempenho ao armazenamento de grandes volumes de informações, foram criados os bancos de dados NoSQL, também conhecidos como não-relacionais. Além de alta performance, bancos NoSQL tem como vantagens o alto *throughput* de dados, alto poder de escalabilidade e maior simplicidade no projeto de esquema de dados.

Um exemplo de banco de dados NoSQL a ser considerado chama-se HBase. Baseado em colunas e que permite armazenamento em chave-valor, o HBase define um mapa multidimensional, ordenado, distribuído, persistente e esparso, indexado por uma chave de linha, chave de coluna e um *timestamp*. Du *et al.* (2015) demonstraram que, ao contrário dos sistemas gerenciadores de bancos de dados relacionais tradicionais, o HBase organiza os dados em um cubo de três dimensões, que representam respectivamente a chave da linha, o nome da família da coluna, o nome da coluna e a versão. Células de dados com várias versões diferenciadas por *timestamps* são armazenados na terceira dimensão. O HBase fornece acesso aleatório rápido de todas as linhas com a condição de uma única chave de linha. Este tem como fundamento ser um banco de dados distribuído, tendo como base o Apache Hadoop, o qual é caracterizado como uma estrutura de software de código aberto escrito em Java para armazenamento e processamento distribuído de conjuntos de dados muito grandes em clusters de computadores construídos a partir de hardware de baixo custo (LOHAR *et al.*, 2016). Sendo um banco de dados distribuído, capaz de suportar grandes volumes de dados, recomenda dividir as tabelas em diferentes regiões antes do início do processo de carregamento. Isso exige dividir a tabela em diferentes regiões entre chaves de divisão explícitas e, em seguida, mover as regiões entre os servidores da região localizados em diferentes nós. Ao fazer isso, torna-se possível aproveitar ao máximo o cluster do Hadoop para distribuir o processo de carregamento e, portanto, acelerá-lo (AZQUETA-ALZÚAZ *et al.*, 2017).

## 2.4 Processamento de imagens aéreas

Após a coleta das imagens é necessário iniciar o processo de extração das características dos objetos (animais) e separação do *background* do restante. Delimitação das Regiões de Interesse (*Region of Interest* - ROI), detecção de objetos e localização na imagem. Esta seção define, também, conceitos necessários para obter os resultados através das redes neurais convolucionais.

### 2.4.1 Conceitos básicos

Geralmente, os métodos de correspondência de imagens são baseados em três etapas principais, como segue:

**Detecção de recursos:** Os recursos de imagem são assim selecionados: imagens (uma imagem de referência e uma imagem de entrada), cantos e regiões.

**Descrição do recurso:** Várias propriedades de imagem, como gradientes, textura e bordas de nível de cinza são usados para fornecer representações robustas das características locais extraídas para caracterizá-los e combiná-los.

**Correspondência de recursos:** A correspondência entre os dados extraídos dos recursos de duas imagens é estabelecida usando medidas de similaridade.

Um dos temas mais relevantes no processamento de imagens é a segmentação destas, com o objetivo de separar o objeto que está sendo buscado do *background*. A subtração de *background* é uma técnica comum e amplamente utilizada para gerar uma máscara de primeiro plano, isto é, uma imagem binária contendo os *pixels* (px) pertencentes a objetos em movimento na cena, usando câmeras fixas.

Para extração de características de imagens, destacam-se os seguintes algoritmos:

***Gaussian Mixture Model (GMM):*** Neste modelo, cada pixel em um quadro é modelado na distribuição gaussiana. Primeiro, cada pixel é dividido por sua intensidade no espaço de cores RGB. Este usa um método para modelar cada pixel de fundo por uma mistura de distribuições K Gaussianas ( $K = 3$  a  $5$ ). Os pesos da mistura representam as proporções de tempo que essas cores permanecem na cena. As cores de fundo prováveis são aquelas que ficam mais longas e mais estáticas (RAGOTHAMAN *et al.*, 2016).

***Histogram of Oriented Gradients (HOG):*** Este descritor é comumente aplicado em reconhecimento de objetos e tarefas de detecção humanas (por exemplo, pedestres), sendo computado nas características locais para um ponto de interesse detectado (ou, em alguns casos, densamente amostrado em todos os pontos de imagem). O alvo é centralizado no ponto de interesse e dividido em uma *grid* quadrada. Dentro de cada célula da grade é calculado um histograma de frequência representando a distribuição das orientações da borda dentro da célula. As orientações das bordas são calculadas e quantificadas. As contagens de histograma são concatenadas para formar um vetor para cada célula, que são novamente concatenadas para formar um

vetor para o alvo (HU; COLLOMOSSE, 2013).

***Scale-Invariant Feature Transform (SIFT)***: Este algoritmo extrai características distintas de imagens locais que são invariantes à escala da imagem, rotação, e são robustos a mudanças no ruído, iluminação e distorção. Em imagens de sensoriamento remoto se pode observar uma cena de diferentes pontos de vista, gerando imagens com distorções geométricas locais significativas. Para duas regiões locais conjugadas em duas imagens diferentes, a quantidade dessas distorções aumenta com a distância das principais características dessas imagens, fazendo necessária a correção por meio de um algoritmo. Outro aspecto é a aproximação de descritores locais em sequência de imagens obtidas a partir de VANT, para a formação de uma imagem única (ortomosaico) (SEDAGHAT; EBADI, 2015).

#### **2.4.2 Visão computacional**

De acordo com Ferreira (2017), o objetivo da visão computacional é extrair informação útil das imagens. Ajudando a tomar decisões sobre objetos físicos e cenas através de imagens. Para tomar decisões sobre objetos reais é quase sempre necessário construir alguma descrição ou modelo deles a partir das imagens. Devido a esse fato, muitos pesquisadores definem o objetivo da visão computacional como a construção da descrição de cenas obtidas das imagens. A visão computacional veio popularizando nos últimos anos por meio de bibliotecas de software disponíveis por grandes empresas e facilidades de acesso a hardwares de alta velocidade e custo mais baixo, impulsionando um mercado restrito na maioria das vezes a governos e gigantes fabricantes de softwares (SAMUELSSON, 2012).

Complementarmente, o rastreamento de vídeo é, em geral, uma tarefa bastante desafiadora e refere-se ao processo de estimar, ao longo do tempo, a localização de um ou vários objetos em um fluxo de vídeo capturado por uma câmera. Encontra-se em uma grande variedade de produtos comerciais que aliviam o operador de tarefas de concentração demoradas e desafiadoras, como reconhecimento de face, controle de tráfego, aplicação de interação humana, imagens médicas, segurança e vigilância (SAMUELSSON, 2012). O rastreamento de vídeo nos submete a desafios pelas diversas tipos de imagens e variações do alvo a ser rastreado. É preciso realizar tratamentos anteriores em imagens, como remoção de fundo e a delimitação de bordas do alvo.



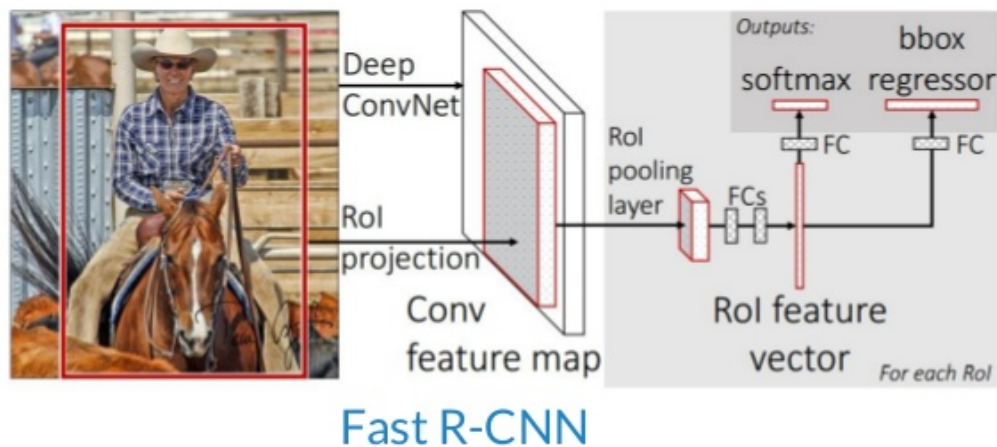
#### *2.4.2.1 Algoritmos de localização de objetos em imagens*

Os algoritmos de detecção e localização de objetos, são capazes de reconhecer um objeto numa imagem através de treinamento prévio com outros objetos alvo a serem buscados. Na sequência são descritos dois dos algoritmos mais utilizados neste processo.

**Fast Region-based Convolutional Network (Fast R-CNN):** Este algoritmo leva como entrada uma imagem inteira e um conjunto de objetos propostos. A rede primeiro processa a imagem inteira com várias camadas convolucionais e de *pool* máximo para produzir um mapa de atributos. Então, para cada objeto, uma camada de agrupamento da ROI extrai um vetor de características de comprimento fixo do mapa de atributos. De acordo com o fluxo mostrado na Figura 2, cada vetor de características é alimentado em uma sequência de camadas totalmente conectadas que finalmente se ramificam para duas camadas de saída: uma que produz estimativas de probabilidade de *softmax* em classes de objetos, mais uma classe de “fundo”, e, por fim, a camada que produz quatro números de valor real para cada uma das classes de objetos. Cada conjunto de 4 valores codifica posições refinadas de caixa (*bounding box*) delimitadora para cada uma das classes.

Figura 2 – Esquema de funcionamento do algoritmo FAST R-CNN.

## Object Detection Network



15

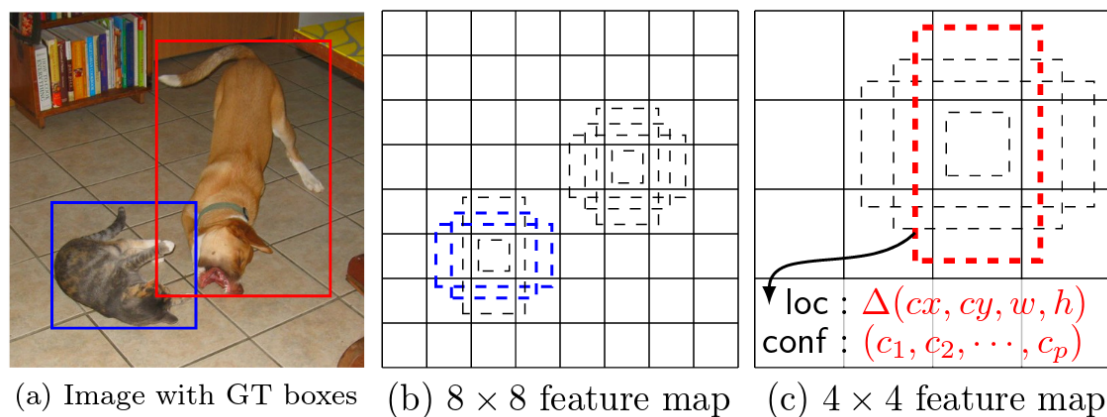
Fonte: Girshick (2015)

**Faster Region-based Convolutional Network (Faster R-CNN):** Este algoritmo representa uma melhoria em relação ao anterior Fast R-CNN em relação ao custo computacional. Na fase de objetos propostos consiste em dois módulos: primeiro, chamado de Rede Regional de Propostas (RPN), uma rede totalmente convolucional para gerar propostas de objetos que serão introduzidos na segunda fase. A segunda é o detector Fast R-CNN cuja finalidade é refinar as propostas. Portanto, compartilhando as mesmas camadas convolucionais para o detector RPN e Fast R-CNN até

as suas próprias camadas totalmente conectadas. Desta forma, a imagem só passa pela CNN uma vez para produzir e refinar as propostas de objetos. Com as camadas convolucionais de compartilhamento, é possível usar uma rede muito profunda para gerar objetos de alta qualidade.

**Single Shot MultiBox Detector (SSD):** Esta abordagem é baseada em uma rede convolutiva *feed-forward* que produz uma coleção de tamanho fixo de *bounding box* e seus scores para a presença de classe de objeto nessas caixas. Após, ocorre uma supressão das camadas para produzir as detecções finais. Como mostra a Figura 3, as primeiras camadas de rede são baseadas em uma arquitetura padrão usada para classificação de imagem de alta qualidade (truncada antes de qualquer camada de classificação). Em seguida, foi adicionada a estrutura auxiliar à rede para produzir detecções com os mapas de características de escala múltipla para detecção de objeto. Essas camadas diminuem gradualmente de tamanho e permitem previsões de detecções em escalas múltiplas. O modelo convolutional para prever detecções é diferente para cada camada de recurso que operam em um único mapa de atributos de escala (LIU *et al.*, 2016).

Figura 3 – Esquema de funcionamento do algoritmo SSD



Fonte: Liu *et al.* (2016)

Na sequência são destacadas técnicas de aprendizado de máquina, com potencial de uso para o reconhecimento de padrões em imagens de forma semiautônoma, ou seja, com a diminuição de etapas de pré e pós-processamento, tornando o método viável para a execução automatizada da identificação de animais.

### 2.4.3 Aprendizado de máquina

Para construir um modelo de treinamento, é preciso ensinar a máquina com outras imagens pré processadas que contenham os padrões desejados. Para tal tarefa existe uma área da inteligência artificial chamada aprendizado de máquina, onde se divide em supervisionado e não supervisionado. O primeiro, que tem como conceito o processo de aprender um conjunto de regras a partir de instâncias ou exemplos de um conjunto de treinamento com a intervenção humana. O aprendizado de máquina não supervisionada é aquele onde algoritmos são treinados com dados de entrada e tentam criar *clusters* (agrupamentos) de forma automática – sem intervenção humana – classificando dados de acordo com o aprendizado.

Também pode ser definido como a criação de um classificador que pode ser generalizado para novas instâncias, não presentes no conjunto de treinamento. Um classificador é um modelo que, após treinado pode ser utilizado para definir classes a instâncias de testes, cujas classes são desconhecidas, utilizando a informação dos seus atributos (FERREIRA, 2017).

Existem diversas técnicas de aprendizado de máquina para reconhecimento de padrões, algumas com necessidades de realização de um pré-processamento maior, outras com características específicas para voz e imagens, e outras com um custo de processamento computacional maior que outras. A partir daqui discutiremos as mais indicadas para processamento em imagens.

#### 2.4.3.1 Redes neurais artificiais

As Redes Neurais Artificiais (RNA) tem como motivação o aprendizado humano, onde o cérebro é estruturado como um computador paralelo, complexo e não-linear, com capacidade de organizar estruturas básicas (neurônios), de tal maneira que possa realizar cálculos de forma rápida e paralela. O processo de aprendizado é definido como a nova conexão entre os neurônios e suas modificações. Com base nesses princípios surgem algo artificial com o objetivo de aproximar aos neurônios humanos chamadas RNA.

A transmissão do sinal de um neurônio a outro no cérebro é um processo químico complexo, no qual substâncias específicas são liberadas pelo neurônio transmissor. O efeito é um aumento ou uma queda no potencial elétrico no corpo da célula receptora. Se este potencial alcançar o limite de ativação da célula, um pulso ou uma ação de potência e duração fixa é enviado para outros neurônios. Diz-se então que o neurônio está ativo. De

forma semelhante à sua contrapartida biológica, uma RNA possui um sistema de neurônios artificiais (também conhecidas como unidades de processamento ou simplesmente unidades). Cada unidade realiza uma computação baseada nas demais unidades com as quais está conectada (BEZERRA, 2016).

Um neurônio artificial é composto por três elementos básicos: conjunto de sinapses definidas com um peso, mais especificamente, um sinal de entrada conectado ao neurônio e multiplicado pelo peso da sinapse.

O segundo elemento é um somador responsável pela adição do resultado da multiplicação dos sinais de entrada pelas sinapses do neurônio.

Os neurônios de uma RNA são organizados em camadas, com conexões entre neurônios de camadas consecutivas. As conexões entre unidades são ponderadas pelos pesos. A RNA mais simples possível contém uma única camada, composta por um único neurônio.

Redes dessa natureza são bastante limitadas, porque resolvem apenas processos de decisão binários e linearmente separáveis. No entanto, é possível construir redes mais complexas por meio de um procedimento de composição de blocos de computação mais simples organizados em camadas. Ao final do processo, o neurônio tem uma função de ativação, que define o resultado, contendo a probabilidade da classificação correta.

Os principais tipos de RNA são: *perceptron*, *multilayer perceptron* e *backpropagation*. O *perceptron* aprende conceitos, ele pode aprender a responder com verdadeiro (1) ou falso (0) pelas entradas apresentadas a ele, sendo treinado repetidamente a partir dos exemplos que lhe são apresentados.

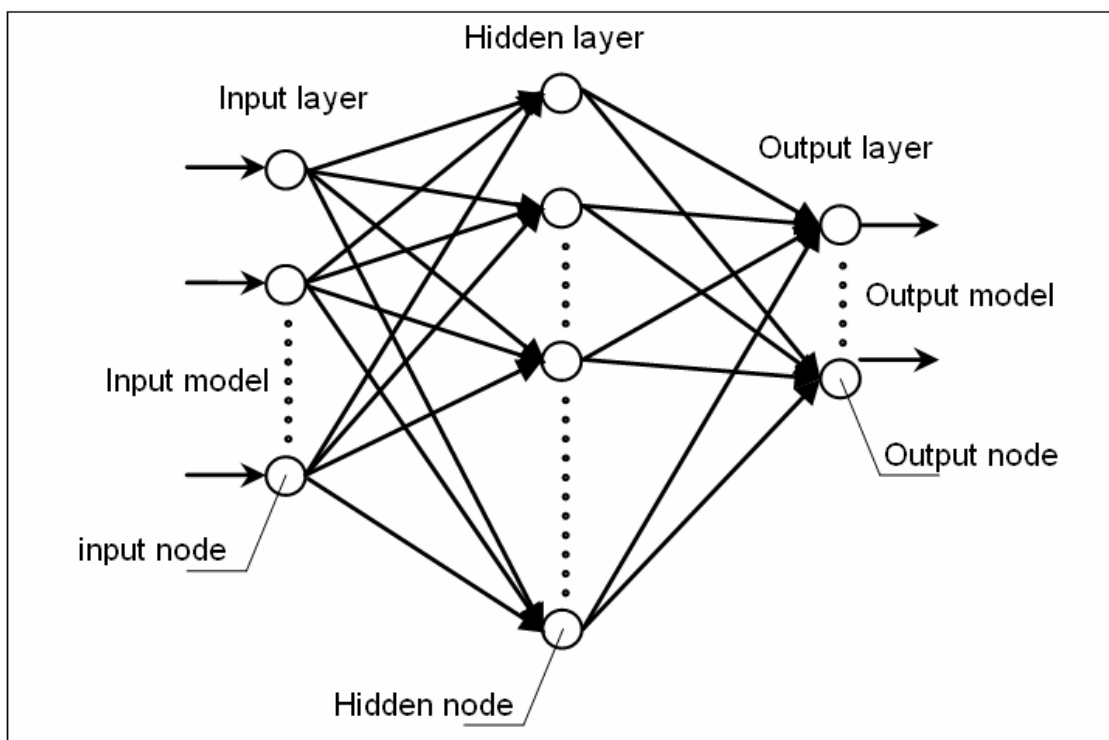
O *perceptron* é uma rede neural cujos pesos e inclinações podem ser treinados para produzir um vetor alvo que quando apresentamos tem que corresponder ao vetor de entrada. Este tipo de RNA tem limitações por trabalharem apenas com valores binários e treinar somente vetores linearmente separados (ARAÚJO *et al.*, 2015).

Um outro tipo de RNA são as *multilayer perceptron neural network* (MLP), que permite várias camadas na rede, contendo vantagens, como a distribuição do conjunto de dados de treinamento não depende de pré-suposições – permitindo a classificação não supervisionada, nenhuma decisão precisa ser definida em relação à importância relativa das várias medições de entrada e a maioria das medições de entrada é selecionada com base no ajuste do peso durante o processo de treinamento.

Camadas de entrada, camadas ocultas e camadas de saída são três composições principais que constroem as RNA MLP (Figura 4). As camadas de entrada são entendidas

como fatores que afetam a definição dos pesos iniciais dos neurônios, as camadas de saída são vistas como os resultados classificados, e as camadas ocultas são as camadas classificadoras para transformar entradas em saídas. São treinadas em duas etapas principais: as entradas são propagadas para frente através das camadas ocultas para produzir os valores de saída e, em seguida, os valores de saída são comparados aos pré-valores para estimar a diferença: a conexão e os pesos são ajustados para otimizar os melhores resultados com a menor diferença (PHAM *et al.*, 2017).

Figura 4 – RNA multilayer perceptron



Fonte: Derras, Bekkouche e Zendagui (2010)

Outro tipo de RNA, chamada *backpropagation*, tem como objetivo calcular a distribuição de erro de cada neurônio. Classificada como treinamento supervisionado, isso significa que os dados precisam ser rotulados. Esta ainda permite encontrar padrões não lineares nos dados rotulados através de multicamadas.

Pradeep, Srinivasan e Himavathi (2011) propuseram um sistema de reconhecimento de texto escrito à mão utilizando imagens como dados de entrada da RNA *backpropagation*. No entanto, visto o uso da RNA com imagens, um conjunto de procedimentos de pré-processamento e extração de características foram necessários para classificar as imagens. O processo de binarização adotado converte uma imagem em escala de cinza em uma imagem binária, usando a técnica de limiar global. Detecção de bordas na ima-

gem binarizada usando técnica de sobel, dilatação da imagem e preenchimento de áreas cortadas são as operações realizadas para produzir a imagem pré-processada adequada para segmentação. Outra etapa do pré-processamento foi decompor as sub-imagens para cada caracter. Já na extração de características, o correto funcionamento melhora a taxa de reconhecimento e reduz a classificação incorreta. Cada imagem de tamanho 90x60 pixels foi dividida em 54 zonas iguais, cada uma com tamanho 10x10 pixels. Os recursos são extraídos de cada pixel da zona movendo ao longo das diagonais de seus respectivos 10x10 pixels. Cada zona tem 19 linhas diagonais e os pixels de primeiro plano presentes ao longo de cada linha diagonal são somados para obter uma única sub-característica. Assim, 19 sub-características são obtidas a partir de cada zona. Estes valores de 19 sub-características são calculados para formar um único valor de característica e colocados na zona correspondente. Este procedimento é repetido sequencialmente para todas as zonas. Pode haver algumas zonas cujas diagonais estão vazias no primeiro plano. Os valores de recurso correspondentes a essas zonas são zero. Finalmente, 54 recursos são extraídos para cada imagem. Em particular, foi possível identificar um longo trabalho feito no pré-processamento das imagens para atingir uma alta acurácia, permitindo concluir que a extração de características com rede neural *backpropagation* foi capaz de produzir uma alta precisão de reconhecimento de 96,52% com 54 características e 97,84% com 69 características.

Outro trabalho que cabe destacar é o de Wang *et al.* (2015), o qual explorou a classificação de frutas por meio de imagens utilizando RNA *backpropagation*, no entanto, foram utilizados alguns passos para realizar o pré-processamento: *discrete wavelet transform* (DWT), técnica para trabalhar com escalas e posição das imagens; *wavelet-entropy*, usada para decompor as imagens e descobrir 30 características; e *Principal component analysis* (PCA), para remover as características descobertas em excesso, com o objetivo de manter o desempenho da classificação. O treinamento foi feito por uma RNA *backpropagation*, chegando a uma acurácia de 89,5%.

Considerando os trabalhos relatados, a RNA apresentou uma acurácia acima de 80% na classificação de imagens, no entanto foram necessárias algumas etapas de pré-processamento para extrair um conjunto pequeno de características relevantes.

#### 2.4.3.2 Aprendizado profundo

Embora as RNA tradicionais apresentem significativa acurácia para classificação de imagens, mostra-se relevante identificar outro tipo de aprendizado de máquina aplicá-

vel a reconhecimento de imagens, capaz de reduzir o pré-processamento, automatizar o processo de extração de características da imagem e promover uma acurácia maior. Neste sentido tem se destacado a área conhecida como aprendizado profundo (*deep learning*).

O aprendizado profundo é definido como um conjunto de técnicas de aprendizado de máquina que exploram muitas camadas de processamento de informação não linear para extração e transformação supervisionada ou não supervisionada para análise de padrões e classificação. Na visão computacional, a utilização vem crescendo progressivamente, principalmente no reconhecimento de objetos em imagens. Por muitos anos, o reconhecimento de objetos em imagens ficou dependente de técnicas como SIFT e HOG. No entanto, estas técnicas têm dificuldade de obter maior nível de informação nas imagens como bordas e fragmentos de objetos. A partir deste problema o aprendizado profundo visa superar as dificuldades obtendo um nível de detalhamento maior dos dados da imagem (FERREIRA, 2017). Dentro deste contexto, destacam-se as chamadas redes neurais convolucionais.

Redes neurais convolucionais (*Convolutional Neural Networks* - CNN) são arquiteturas biologicamente inspiradas capazes de serem treinadas e aprenderem representações invariantes à escala, translação, rotação e transformações afins. As CNN são projetadas para uso com dados em duas dimensões tornando-as uma boa candidata para a solução de problemas envolvendo reconhecimento de imagens. Por definição, uma arquitetura profunda é uma estrutura hierárquica de múltiplas etapas, onde cada etapa é formada por uma rede neuronal de, pelo menos, três camadas, e cada etapa é treinada pelo algoritmo *backpropagation*. Com as fontes em larga escala de dados de treinamento e implementação eficiente em GPU, as CNN recentemente superaram alguns outros métodos convencionais, até mesmo desempenho humano, em muitas tarefas relacionadas à visão, incluindo classificação de imagens, detecção de objetos, rotulagem de cena e reconhecimento facial. Além das tarefas de visão, as CNN também foram aplicadas em outras áreas, como o reconhecimento de fala. A técnica foi verificada como uma classe eficaz de modelos para a compreensão do conteúdo da imagem visual, dando alguns resultados de última geração na classificação da imagem visual e outros problemas relacionados à imagem (HU *et al.*, 2015). Na arquitetura, a própria rede teria que detectar as dependências existentes na estrutura espacial da distribuição subjacente às imagens de entrada. Além disso, devido à conectividade muito grande, esse tipo de arquitetura sofre da denominada maldição da dimensionalidade (*curse of dimensionality*) e portanto não é adequado a imagens de alta resolução, por conta do alto potencial de sobreajuste dos dados. Isso



sem considerar o tempo para computar as pré-ativações de todas as unidades em cada camada (BEZERRA, 2016).

Nas CNN, as características de uma imagem são detectadas através de conjuntos distintos, denominados mapa de características (*feature map*), conhecido como *kernel*. As unidades contidas em um mapa de características são únicas na medida em que cada uma delas está ligada a um conjunto de unidades diferente na camada anterior. Além disso, todas as unidades de um mapa compartilham os mesmos parâmetros. O resultado disso é que essas unidades dentro de um mapa servem como detectores de uma mesma característica, mas cada uma delas está conectada a uma região diferente da imagem. Portanto, em uma CNN, uma camada oculta é segmentada em diversos mapas de características, em que cada unidade de um mapa tem o objetivo realizar a mesma operação sobre a imagem de entrada, com cada unidade aplicando essa operação em uma região específica dessa imagem.

Uma unidade de um mapa de característica, realiza uma operação chamada convolução. Esta ação determina a realização do produto Hadamard entre a matriz original e outra matriz, denominada núcleo de convolução (*convolution kernel*). A matriz deve ser definida para encontrar níveis maiores ou menores de características. Quanto maior o kernel maior o custo de computacional necessário para extração das características, no entanto, obtém-se uma acurácia maior.

Outra operação importante utilizada em uma CNN é a subamostragem. Em processamento de imagens, a subamostragem de uma imagem envolve reduzir a sua resolução, sem, no entanto, alterar significativamente o seu aspecto. No contexto de uma CNN, a subamostragem reduz a dimensionalidade de um mapa de característica fornecido como entrada e produz outro mapa de característica. Há várias formas de subamostragem aplicáveis a um mapa de característica: selecionar o valor máximo (*max pooling*), a média (*average pooling*) ou a norma do conjunto (*L2- pooling*).

Na Figura 5 é exposta uma imagem no formato original e o comportamento após aplicação da convolução na Figura 6. Note que na camada convolucional é possível destacar as bordas da imagem através do relevo, cor específica, ou apenas borrar o ruído indesejado.

No exemplo da Figura 7, demonstra-se uma imagem 6x6 com *kernel* tamanho 3x3. O produto da soma dos pixels da imagem a cada matriz 3x3, conforme o exemplo 429 (18+51+121+35+204). Outro objetivo da subamostragem é a localização exata de característica da imagem, tornando invariável a pequenas translações.

Figura 5 – Imagem antes da aplicação do *kernel* 3x3



Fonte: Autor (2019)

Figura 6 – Imagem após aplicação do *kernel* 3x3



Fonte: Autor (2019)

Figura 7 – Subamostragem

INPUT IMAGE					
18	54	51	239	244	188
55	121	75	78	95	88
35	24	204	113	109	221
3	154	104	235	25	130
15	253	225	159	78	233
68	85	180	214	245	0

WEIGHT		
1	0	1
0	1	0
1	0	1

429

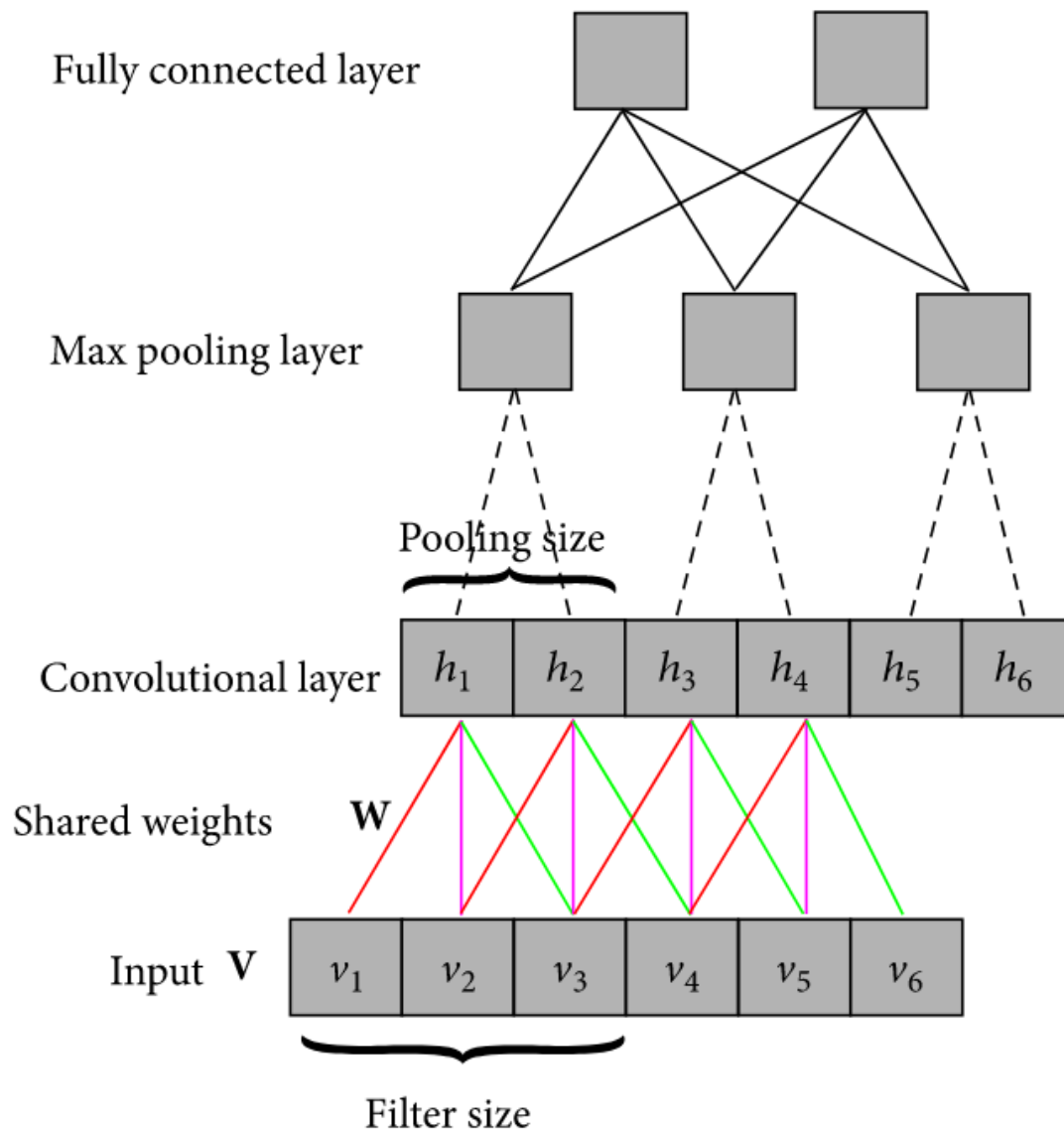
Fonte: Autor (2019)

As camadas das CNN possuem diversos níveis de abstração que dão a possibilidade de segmentação de imagens, eliminando a fase de pré-tratamento destas. Cada ativação oculta é calculada multiplicando uma pequena entrada local pelos pesos. Os pesos são então compartilhados em todo o espaço de entrada, conforme mostrado na Figura 8. Os neurônios que pertencem à mesma camada compartilham os mesmos pesos. O compartilhamento de peso é um princípio crítico nas CNN, pois ajuda a reduzir o número total de parâmetros treináveis e leva a um treinamento mais eficiente e a um modelo mais eficaz. Uma camada convolucional é geralmente seguida por uma camada *max pooling* (HU *et al.*, 2015).

No contexto de arquiteturas de extração de características de imagens, CNN são pré-treinadas com um conjunto de dados de imagens (banco de dados *ImageNet*) para descritores de imagem genérica distintos e podem ser aplicados para extrair características discriminativas de imagens baseadas na teoria de *transfer learning* - neste tipo de CNN, treinada em uma grande imagem da natureza conjunto de dados antes de ser usado como um extrator de recurso em um pequeno conjunto de dados. As características extraídas de CNN pré-treinadas são genéricas e aplicáveis a outros conjuntos de dados (NGUYEN *et al.*, 2018). Destacaram-se na revisão da literatura duas arquiteturas: *Inception* e *Inception-ResNet-V2*. A primeira é capaz de processar recursos espaciais mais ricos e aumentar a diversidade de recursos. Já a segunda arquitetura une o potencial da *Inception*, controlando a degradação através de conexões de atalho que desviam seções mais iniciais para redes mais profundas, melhorando acurácia. As duas suportam *transfer learning* (ZHAO *et al.*, 2018).

O algoritmo *Faster R-CNN* foi usado para classificar padrões de imagens de hu-

Figura 8 – Arquitetura de uma CNN



Fonte: Derras, Bekkouche e Zendagui (2010)

manos em estado correndo, caminhando e correndo muito rápido a partir de vídeos feitos por smartphone com resolução 1920 x 1080 px (YANG; GENG, 2018). O conjunto de dados (*dataset*) foi montado convertendo seus *bouding box* em XML para treinamento. O autor conseguiu 3000 imagens randomicamente para treinamento, *crops* de tamanhos de 200 a 400 px, e obteve uma acurácia de 98,9% nas classes correndo e caminhando e correndo muito rápido 97,7%.

O algoritmo SSD, com imagens coletadas por meio de VANT, foi usado para identificar isoladores em linhas de transmissão de energia (MIAO *et al.*, 2019). Foi coletado um banco de dados de 6700 imagens fazendo *crops* de tamanhos de 150 a 400 px, sendo destaque que a quantidade não foi suficiente para os resultados. Obteve-se acurácia de 93,75% e 85,29% nas duas classificações de linhas previstas. Com estas variações de algoritmos percebe-se a necessidade de uma quantidade de imagens acima de 3000 exemplos para treinamento. Além disso, identificou-se que *Faster R-CNN* tende a produzir melhores resultados com tamanhos de imagens diferentes e quantidades menores de imagens.

Complementarmente, é uma biblioteca de código aberto escrita pelo Google. Este framework obteve imediatamente um grande sucesso na comunidade de Aprendizado de Máquina e em mais de 3 anos também teve muito apoio e desenvolvimento pelo próprio Google, mais por muitos projetos da comunidade, desenvolvidos em qualquer área de *Deep Learning*. A peculiaridade do TensorFlow é seu fluxo de trabalho, feito por gráficos de fluxo de dados. Onde os nós representam operações matemáticas, as arestas representam as matrizes multidimensionais entre elas. Em maio de 2016, o Google revelou que utilizou o TensorFlow no projeto AlphaGo, com um hardware especial dedicado a melhor performances da biblioteca. Para alcançar rapidamente esses objetivos, o Google dedicou uma atenção especial à experiência do usuário do TensorFlow, que chega com suporte a uma comunidade GitHub bem desenvolvida, a chave dessa rápida melhoria. No próximo capítulo são sintetizados os trabalhos correlatos.

### 3 TRABALHOS RELACIONADOS

Neste capítulo é apresentada uma revisão a respeito de trabalhos identificados na literatura que têm relação com reconhecimento de padrões de animais em imagens através de VANT, complementados por outros que tratam de armazenamento de imagens sob a perspectiva de performance e escalabilidade.

#### 3.1 Reconhecimento de animais em imagens

Andrew, Greatwood e Burghardt (2017) usaram redes neurais convolucionais para reconhecer de forma individual, através de imagens, a raça Holstein Friesian do gado a campo. Inicialmente, ele fixou uma câmera acima dos animais em uma área restrita onde os animais passavam pelo espaço e registrou 900 imagens. Após treinar o classificador, capturou vídeos aéreos por meio de *drones* a campo para utilizar como validação dos testes. Os autores concluíram que para a identificação em particular, as arquiteturas baseadas em convoluções são bem adequadas para aprender e distinguir as propriedades de padrão e estrutura dorsal únicas exibidas individualmente pelas espécies. Importante ressaltar que este processo pode ocorrer de forma não invasiva, em contraste com a maioria das estruturas de identificação existentes girando em torno de um equipamento instalado no animal.

Gemert *et al.* (2014) objetivaram monitorar de forma automática a distribuição de animais em uma área, com a justificativa de que a tarefa de controlar a extinção de certas espécies de forma manual requer muito trabalho humano e tempo. Como os autores não tinham dados disponíveis, capturaram vídeos através de *drones* e implementaram três classificadores baseados em diferentes técnicas: *Support Vector Machines* (SVM), *Deformable Part Model* (DPM) e *Color DPM*. Os resultados mostraram que o desempenho para detecção em imagens de objetos humanos não se traduz diretamente em imagens de *drone*. De acordo com a literatura, a *color DPM* deveria superar o DPM padrão, que por sua vez deveria superar o SVM. No entanto, os resultados foram exatamente opostos.

Chamoso *et al.* (2014) focaram na obtenção de varreduras de grandes áreas do sistema de pecuária. A contagem e monitoramento de espécies animais podem ser realizadas com gravações de vídeo retiradas de *drones*. Além disso, o sistema acompanha o número de animais detectados através da análise das imagens coletadas com as câmeras de *drones* em tempo real. Testes foram realizados para avaliar este sistema e obtiveram

os seguintes resultados preliminares: a partir de 70 *frames* analisados, 57 foram corretos, 2 além da conta e 11 não contados. Animais contados foram, acima 24, 309 corretos e 99 não contados em um total de 432.

Rey (2016) buscou identificar a quantidade de animais em savanas semiáridas, justificando as ameaças recebidas pelas mudanças no equilíbrio frágil entre chuvas, incêndios e pressão de pastagem exercida por animais selvagens ou gado. Para evitar a invasão dos arbustos e o declínio da grama perene, os administradores de terras devem prestar atenção para manter a quantidade de gado e vida selvagem em equilíbrio com a disponibilidade de grama. Em grandes fazendas e parques de conservação, estimar as populações de animais é, portanto, um importante aspecto de gestão. Métodos tradicionais de recenseamento animal - tais como contagens de transecto de um helicóptero, ou marca / recaptura - são muito caras e trabalhosas para serem conduzidas regularmente. O autor destacou que VANT aparecem como uma ferramenta para a detecção de animais. Eles podem ser facilmente implantados, por um custo menor e maior segurança. O aspecto negativo é interpretar visualmente o grande número de imagens de alta resolução que elas adquirem. Os recentes avanços nas técnicas de aprendizado de máquina podem permitir a automação da detecção de animais nessas imagens aéreas. Neste contexto, buscou implementar diferentes algoritmos, a fim de investigar a viabilidade e os benefícios potenciais da combinação de aprendizado de máquina e VANT para a detecção de animais. As técnicas de aprendizado de máquina envolvidas foram *Bags of visual Words* (BoW), SVM e aprendizado ativo. Os resultados foram promissores mostram que taxas de recuperação na faixa de 60 a 80% são possíveis, se uma baixa precisão (5 a 20%) for aceita.

### 3.2 Armazenamento espaço-temporal de imagens

Quanto ao armazenamento das imagens, Ferreira e Oliveira (2014) utilizaram um banco de dados NoSQL para armazenar imagens de diagnósticos de câncer de pulmão, com a justificativa de que estas não organizadas em esquemas de banco de dados, e, portanto, descentralizadas e sem correlação entre imagens, dados dos exames e a classificação dos nódulos realizada pelos especialistas, inviabilizariam o gerenciamento e a utilização dos dados.

Ferreira e Oliveira (2014) também destacaram que o uso do banco de dados *MongoDB*, possui fraco acoplamento, logo, o desenvolvimento do banco de dados pode ser aplicado a outras tecnologias e capaz de armazenar grandes volumes de dados. Outro

aspecto relacionado é que o autor segmentou as imagens para visualização, diminuindo assim o tamanho da imagem para armazenamento.

Por outro lado, Xiao e Liu (2011) utilizaram o banco de dados NoSQL HBase distribuído, com capacidade de escala e suporte a grandes volumes de dados, para armazenamento de imagens e metadados oriundos de sensoriamento remoto em uma aplicação WEB. Em especial demonstraram que com o armazenamento massivo de imagens no HBase funciona de forma mais efetiva do que nas alternativas avaliadas, demonstrando estar adaptado às aplicações que usam imagens.

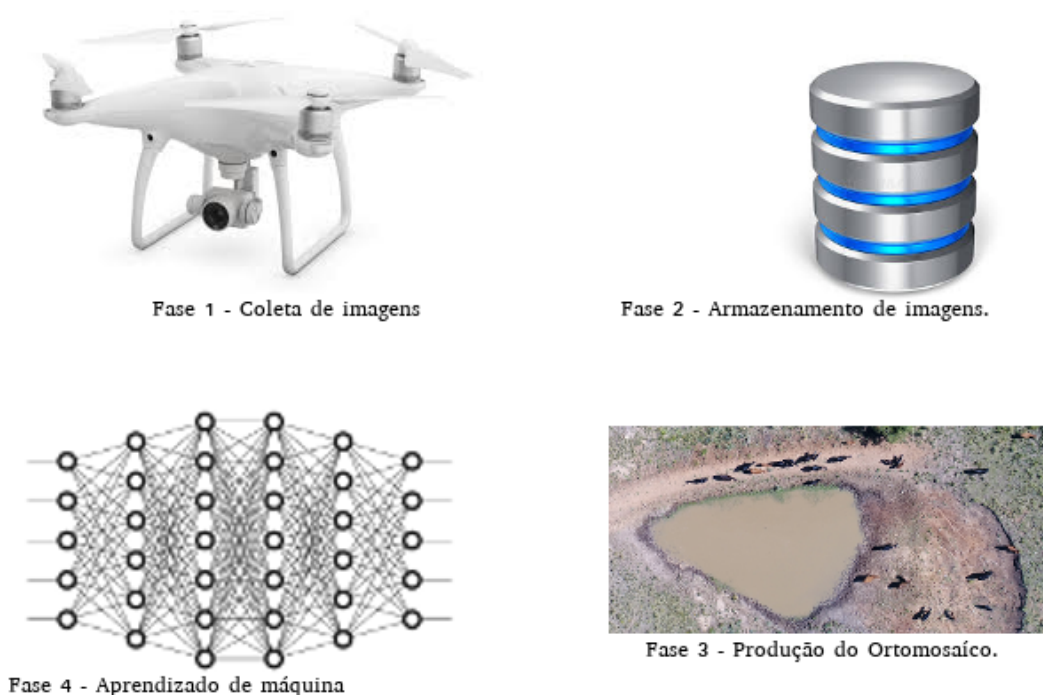
Liu *et al.* (2013) apresentaram um método de armazenamento de imagens *raster* baseado no HBase, envolvendo a divisão das imagens em fragmentos e distribuição entre nós disponíveis no banco de dados. Além disso, devido à alta escalabilidade do Hadoop, mostrou ser fácil adicionar novos nós ao *cluster*, o que geralmente é difícil de maneira tradicional. Foi descoberto também, que as velocidades de importação e o processamento de dados melhora de acordo com o número de *clusters*. Isso significa que mais nós trariam mais capacidade e maior poder de computação. Isso mostra que o HBase é adequado para armazenamento e processamento massivo de dados de imagens, considerando que o Hadoop é projetado para ser executado em um *cluster* grande com milhares de nós.



## 4 MÉTODO PROPOSTO

Este capítulo tem como objetivo descrever o método proposto no trabalho. Este método de identificação de animais a campo tem como base as “tecnologias de apoio ao manejo extensivo de rebanhos em sistemas de produção baseados em conceitos de Pecuária de Precisão”. Ele é dividido em quatro etapas (Figura 9), detalhadas nas seções deste capítulo.

Figura 9 – Passo a passo do método proposto



Fonte: Autor (2019)

### 4.1 Primeira Etapa

Na primeira etapa ocorre o processo de coleta por aerolevantamento, o qual envolve a definição de um plano de voo em uma área que contenha animais e seja acessível por um VANT em dias sem intempérie, já que há necessidade do registro de imagens. A coleta de imagens do espectro visível – parcela do espectro eletromagnético na qual a radiação é composta por fótons capazes de sensibilizar o olho humano – é realizada por meio de uma câmera digital padrão RGB embarcada no VANT, posicionada de forma perpendicular ao veículo.

O tempo entre a captura de cada uma das imagens depende diretamente da velocidade de voo, pois para voos em maior velocidade a câmera deverá obter as imagens com maior frequência. A altitude do voo é determinante para que as técnicas de reconhecimento de objetos em imagens possam identificar os animais à campo. Aquela também influencia na área de abrangência no registro de imagens pelo VANT, já que há uma limitação na bateria deste tipo de equipamento.

## **4.2 Segunda Etapa**

Na segunda etapa se dá o processo de persistência, por meio do qual a sequência de imagens coletadas é armazenada em um sistema gerenciador de banco de dados espacial a fim de manter a temporalidade das imagens para análises futuras e extração de padrões.

## **4.3 Terceira Etapa**

A terceira etapa do processo inicia pelo acesso às imagens armazenadas na base de dados espaço-temporal com metadados associados, um ortomosaico georreferenciado, tendo como objetivo aumentar a eficiência minimizando a duplicidade na contagem causada pela potencial movimentação dos animais durante o aerolevanteamento, referente à área de interesse delimitada por um polígono, aplicando o algoritmo SIFT sobre o conjunto de imagens com 60% de sobreposição, é criado a partir das imagens.

Com o ortomosaico concluído, anotações de polígonos de animais, por meio da definição de Regiões de Interesse (*Regions of Interest* - ROI) são rotulados animais, por classe, suficientes para a primeira etapa de treinamento do classificador baseado em técnica supervisionada.

## **4.4 Quarta Etapa**

Como quarta e última etapa do processo, as imagens rotuladas na etapa anterior são submetidas a um classificador pré-treinado, que tem como objetivo melhorar a performance tendo como referência a métrica tempo de treinamento. A partir do resultado do classificador, as imagens não rotuladas na etapa três são classificadas e as métricas acurácia, que sintetiza a relação entre a quantidade de erros e acertos do método na detecção

dos animais, e a Intersection over Union (IoU), a qual avalia a porcentagem da sobreposição da predição feita corretamente sobre o objeto real na imagem, são avaliadas, obtendo como resultado a contagem dos animais de forma automatizada, associada à identificação da posição geográfica aproximada de cada animal.

## 5 METODOLOGIA EXPERIMENTAL

Este capítulo descreve os materiais e métodos utilizados na metodologia experimental adotada para avaliação empírica do método proposto em um cenário real. Na Seção 5.1 são caracterizados a área mapeada, os equipamentos e os procedimentos utilizados nos aerolevantamentos e na transmissão dos dados coletados, o ambiente computacional, bem como as ferramentas de processamentos e armazenamento de imagens adotadas nos experimentos. Na Seção 5.2 são detalhados os experimentos realizados para avaliação do método proposto. Por fim, na Seção 5.3 são apresentadas as métricas escolhidas para avaliar quantitativamente o método.

### 5.1 Materiais e métodos

#### 5.1.1 Área mapeada

A área mapeada pelo VANT foi uma área de 20 ha, disponível na Embrapa Pecuária Sul, situada no município de Bagé, Rio Grande do Sul. Com campo nativo, devido a região, encontram-se animais de diversas raças distribuídos de forma aleatória ao longo da área com relevo suavemente ondulado. Para realização do mapeamento da área especificada através da grametria foram necessárias múltiplas imagens aéreas individuais em sequência, ao longo de 50 linhas de voo com sobreposição de imagens. A sobreposição está relacionada com a porcentagem que uma fotografia se sobrepõem a outra, permitindo o reconhecimento de pontos em comum entre as duas ou mais imagens.

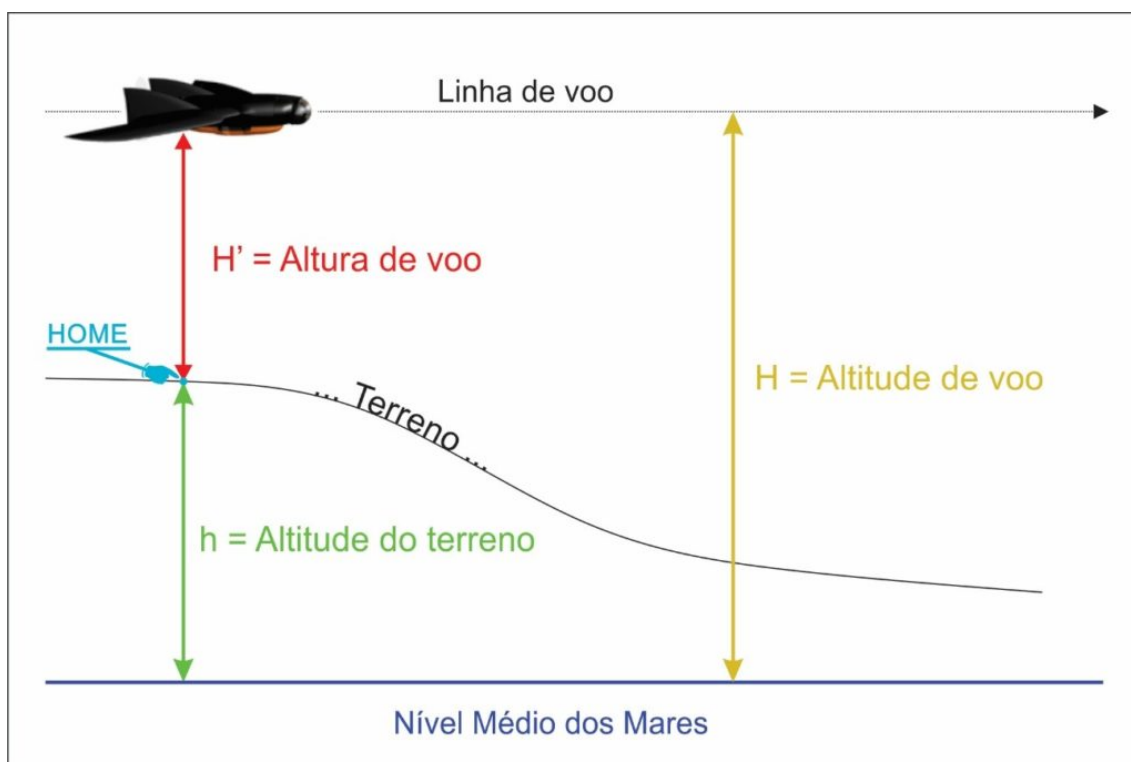
#### 5.1.2 Equipamentos utilizados

Para realização da cobertura da área foi utilizado um VANT multirrotor da marca DJI Phantom 4 com a câmera Survey2 RGB com 16 MegaPixel (4.608 x 3.456 px), *Ground Sample Distance* (GSD) de 4,05 cm/px a 120 m AGL, Sensor Sony Exmor IMX206 de 16 MP (Bayer RGB), velocidade de captura das imagens em JPG de 2 segundos e armazenamento em Micro SD.

### 5.1.3 Imageamento da área

Considerado como fator determinante para a correta coleta das imagens a campo, o GSD - define a resolução espacial das imagens - deve ser planejado com base nas características da câmera e a definição necessária para análise posterior das imagens por algoritmos de aprendizado de máquina. Na Figura 10 é demonstrada a variação do GSD de acordo com a mudança de altitude do voo (H) e terreno.

Figura 10 – Linha de voo e GSD com variação de altitude



Fonte: Autor (2019)

Neste trabalho, com base na câmera do VANT Phantom 4, foi aplicada a equação:

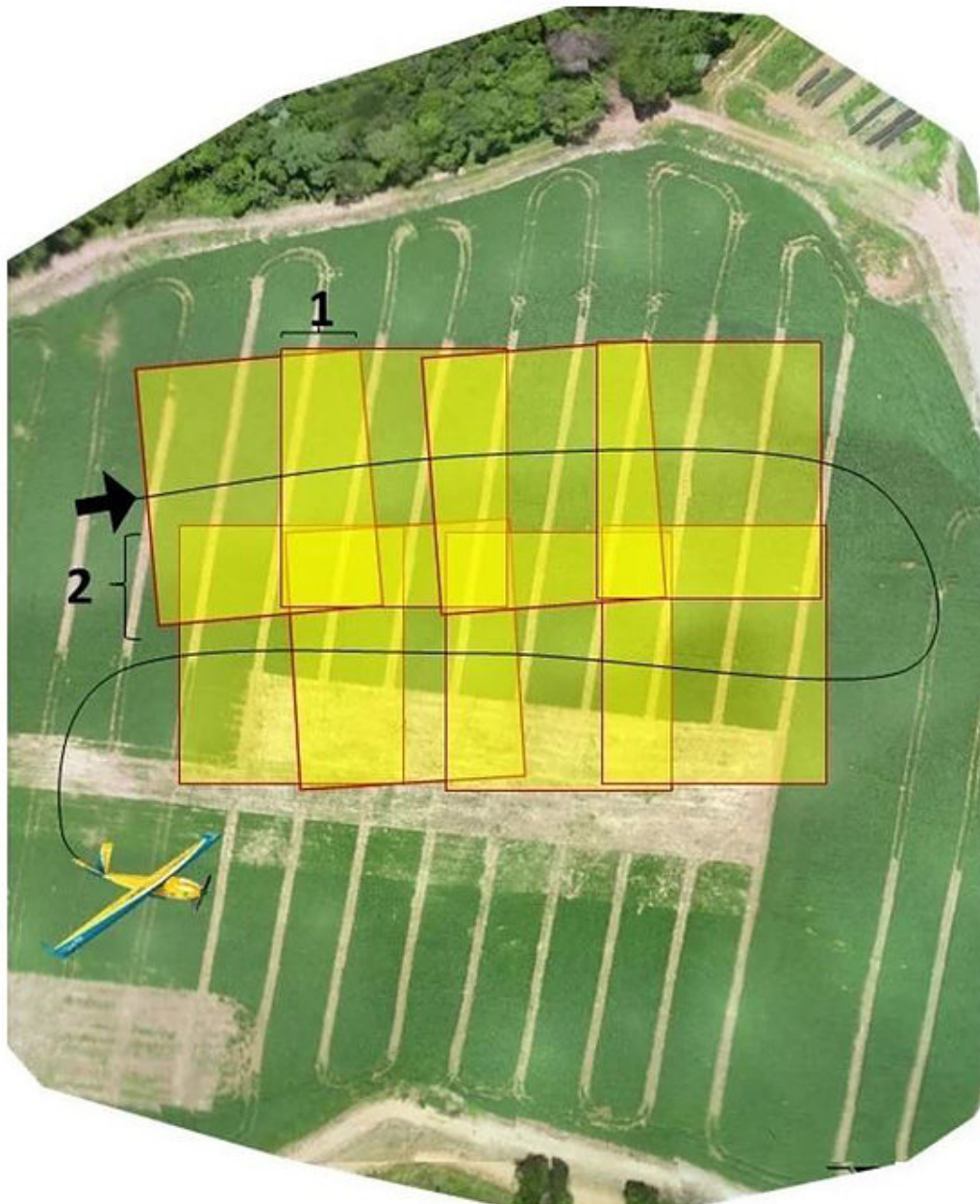
$$\frac{(100.0 * largura\ do\ sensor * altura\ do\ voo)}{(tamanho\ focal * largura\ da\ imagem)} \quad (1)$$

Onde, a largura do sensor é 13,2, altura do voo definida em 100 m, tamanho focal de 8,8, largura da imagem a ser gerada pelo VANT de 4000 px. Como resultado, teve-se um GSD de 3,75 cm.

Foram feitos dois voos - em agosto e novembro de 2018 - em dias distintos com sol, temperatura de 20 graus, vento com cerca de 10 km/h, e variação de horário: 16 h e 12 h, respectivamente.

A área na qual foram coletadas as imagens têm um terreno suavemente ondulado, sem arbustos ou árvores. Foram necessárias 50 linhas de voo de fotografia aérea vertical a 100 m altitude no formato demonstrado na Figura 11, de forma ilustrativa, com seqüências de tomadas de fotos em intervalo de dois segundos.

Figura 11 – Linhas de voo e sobreposição



Fonte: Autor (2019)

Durante o voo, foi realizado um sobrevoo com 60% de sobreposição de imagens para aproveitamento de diversos ângulos e luminosidade da radiação solar.

#### 5.1.4 Transmissão de dados

Após a coleta dos dados e armazenados temporário no cartão SD da câmera, os dados foram transmitidos para o Google Colaboratory (ambiente dos experimentos descrito na próxima subseção) por meio de script.

#### 5.1.5 Ambiente computacional

Notou-se que, para dar suporte às necessidades computacionais de operações paralelas com imagens, deve-se destacar que arquiteturas *single-threading* e superescalares tradicionais vêm dando lugar a abordagens diferentes a fim de aumentar o desempenho e reduzir o consumo de energia. O projeto de processadores com técnicas *multithreading* e múltiplos núcleos de processamento vem sendo consolidado como uma alternativa para melhorar o desempenho computacional. Para a nova geração de processadores com arquiteturas *many-core*, a quantidade elevada de núcleos demanda um novo tipo de projeto baseado em um sistema heterogêneo com diversos núcleos diferentes, memórias cache, redes de interconexões, protocolos de coerência adequados, a fim de obter um alto desempenho no sistema final (ALVES; FREITAS; NAVAUX, 2012).

O uso de arquiteturas computacionais, como sistemas compostos por CPU de múltiplos núcleos e *Graphics Processing Unit* (GPU), capazes de processar paralelamente operações em um conjunto de imagens, deve ser levado em consideração ao propor soluções envolvendo reconhecimento de padrões. No método desenvolvido, utilizou-se a biblioteca de computação numérica chamada TensorFlow, a qual possui suporte a processamento paralelo em CPU com múltiplos cores e também GPU.

Para tanto, o Google disponibiliza uma máquina virtual (VM) com Linux e ambiente de desenvolvimento Python 2 e 3 já preparados para pesquisas voltadas para inteligência artificial com CPU e GPU de forma gratuita, totalmente em *cloud*, chamado Google Colaboratory (Colab). No Colab uma VM com ou sem GPU é acionada por demanda, permitindo avaliar a execução de um programa completo ou em blocos de código. A plataforma pode ser útil na aceleração de processamentos computacionais de projetos de aprendizado de máquina nos quais o fator tempo é determinante. Tem como base o ambiente web Jupyter Notebook, o qual permite acesso remoto sem qualquer tipo de instalação necessária e o compartilhamento de scripts de execução de qualquer linguagem interpretada, principalmente projetos Python, podendo ser mais rápidos que 20 cores

físicos em aplicação de aprendizado profundo, eliminando a etapa de configuração de ambientes de execução para treinamentos naquela aplicação (CARNEIRO *et al.*, 2018).

### 5.1.6 Processamento de imagens

A partir dos dois voos realizados para este trabalho foram gerados ortomosaicos com o software *OpenDroneMap* 0.3. O ambiente disponível para processamento foi um computador com processador Core i7, 4 cores, memória RAM disponível 8GB, sem GPU, disco rígido de 1TB SATA 7.200 RPM e sistema operacional Linux 4.13.0-38 64 bits. O processo demorou cerca de 2 h em cada voo.

O ortomosaico resultante do voo 1 pode ser visto na Figura 12, onde animais encontram-se parados e na posição em pé. É possível identificar a dificuldade no reconhecimento do animal até mesmo por visualização humana, em razão do tamanho dos animais nas imagens.

Neste voo observou-se alguma sombra gerada pelos animais em razão do horário (16 h), conforme Figura 13, ponto a ser observado no treinamento da rede neural para não haver erro na classificação encontrando algum animal duplicado. Outro aspecto relevante, é a posição da câmera do drone, dando a necessidade de rotação por parte do sensor - característica suprimida pelas redes neurais convolucionais.

O ortomosaico gerado do voo 2 pode ser visto na Figura 14, onde animais encontram-se parados também. Neste voo não se observou sombra, em razão do horário do voo, mas é possível verificar árvores e arbustos. No entanto os animais não são vistos próximos.

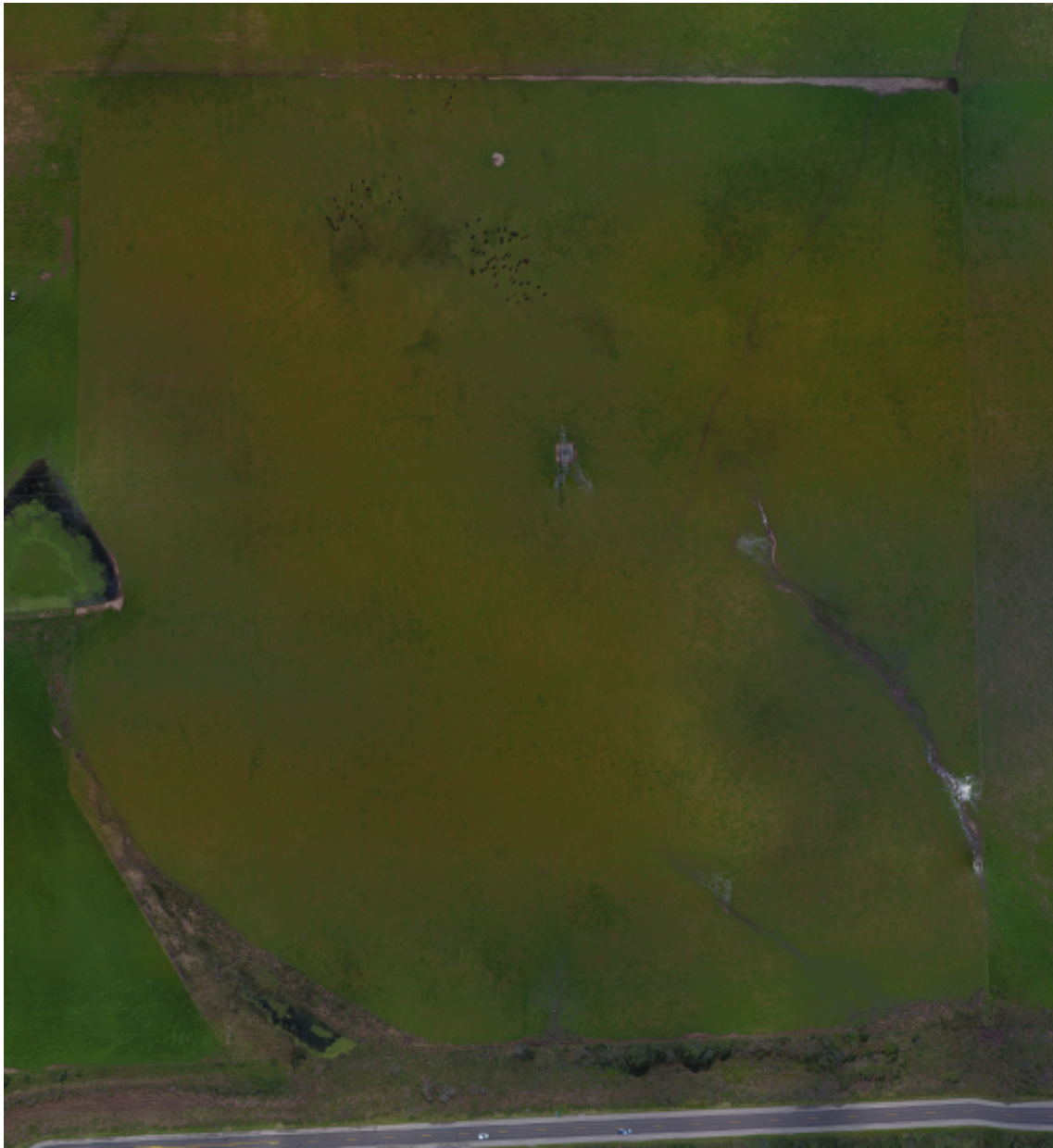
### 5.1.7 Armazenamento de imagens

Com o objetivo de manter o aspecto temporal das imagens e também os metadados para pesquisas das imagens, o armazenamento foi feito no banco de dados NoSQL HBase.

Conforme modelo colunar existente no HBase, a proposta do modelo de dados deste trabalho consta na Figura 15, a linha chave corresponde a cada imagem, na família de coluna 1 constam as colunas nome\_imagem, arquivo binário, tipo de dados, resolução da imagem, latitude e longitude.



Figura 12 – Ortomosaico Voo 1



Fonte: Autor (2019)

## 5.2 Experimentos

Considerando os animais como objetos a serem contados na área a ser imageada, alguma técnica de aprendizado de máquina torna-se imprescindível para identificação deles. Em razão de, o estado da arte da detecção de objetos serem as CNN, por sua capacidade de extração de características em imagens serem superiores em relação a outras técnicas, esta opção foi selecionada.

Não é comum treinar uma CNN com inicializações aleatórias de pesos, pois seria

Figura 13 – Vacas e geração de sombra



Fonte: Autor (2019)

necessário uma grande quantidade de imagens e algumas semanas de treinamento utilizando múltiplas GPU. Com isso, uma prática comum consiste em utilizar os pesos de uma rede já treinada para uma base muito grande, como a ImageNet (DENG *et al.*, 2009) que possui mais de 1 milhão de imagens e 1000 classes. Em seguida, esses pesos podem ser utilizados para inicializar e retreinar uma rede, ou mesmo para a extração de características de imagens (RUSSAKOVSKY *et al.*, 2015).

Portanto, fazendo uso de uma rede neural convolucional já com imagens treinadas anteriormente, inicia-se o processo de transferência de aprendizado, onde a última camada recebe novas classes vaca e vaca deitada, aproveitando o aprendizado anterior, onde espera-se um aumento da métrica tempo de execução no processo de treinamento da rede.

Utilizando a biblioteca TensorFlow com o algoritmo *backpropagation* e as imagens devidamente segmentadas, são dadas as entradas para treinamento na rede com as

Figura 14 – Ortomosaico Voo 2



Fonte: Autor (2019)

Figura 15 – Modelo de dados HBase

Linha Chave	Família de Coluna 1						Família Coluna 2
	Colunas						Colunas
Imagem1	nome_imagem	arquivo	tipo	resolucao	latitude	longitude	nome_imagem

Fonte: Autor (2019)

imagens anteriormente classificadas como terreno e animais, delimitando em 60% das imagens e 40% para testes, tendo como base inicial de treinamento a ImageNet reaproveitando os pesos iniciais e classificações.

Os hiperparâmetros praticados com o algoritmo *Faster R-CNN* arquitetura *Inception Resnet v2* foram: variação de escala e proporção de 0,25 a 3,0, *stride* de tamanho 16, *batch size* 12, dividido em épocas de 600 e 1100, e *threshold* mínimo de 60% e com capacidade máxima de detecção definida para 300 *bounding boxes* por imagem, *learning rate* 0,0002 a cada 200 épocas.

A metodologia adotada nos experimentos conduzidos para avaliar o método proposto foi dividida em seis etapas, conforme mostra a Figura 16:

- Aerolevanteamento da área de interesse;
- Armazenamento das imagens em banco de dados espaciais;
- Criação de um ortomosaico;
- Segmentação dos objetos na imagem;
- Treinamento dos classificadores;
- Teste dos classificadores;
- Validação dos classificadores.

Na primeira etapa são coletadas imagens dos animais a campo na área disponível, a partir de um plano de voo parametrizado com objetivo de coleta de imagens RGB com sobreposição horizontal e vertical de 60% com determinada resolução espacial (GSD): 5 cm por pixel.

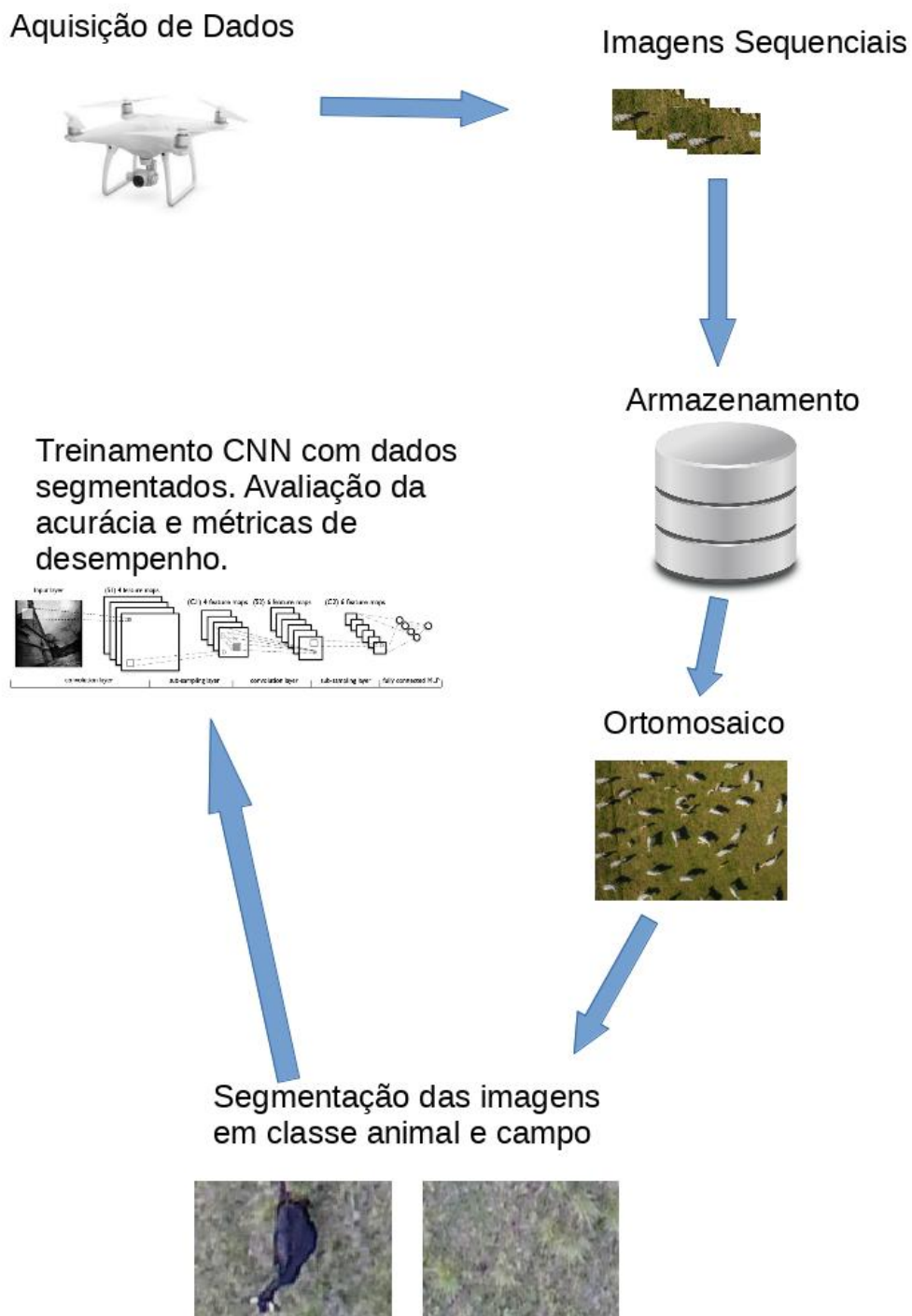
Na segunda etapa são armazenadas imagens com associação de metadados do voo e da área experimental.

Na terceira etapa são unidas as imagens sequenciais a partir de suas bordas, com o objetivo de aumentar a acurácia minimizando a duplicidade na contagem causada pela potencial movimentação dos animais durante o aerolevanteamento.

Na quarta etapa é aplicado um processo de reconhecimento de padrões em imagens, identificado a presença de objetos candidatos à classificação como animal, incluindo a separação do *background* (terreno) dos animais.

Na etapa cinco, o ortomosaico é segmentado em imagens de 400 pixels quadrados, sendo realizado um treinamento com classificador baseado em técnicas amparadas por princípios de visão computacional outro baseado em rede neural convolucional, com um conjunto de treinamento escolhido aleatoriamente, correspondente a 60% das ima-

Figura 16 – Passo a passo do método proposto



Fonte: Autor (2019)

gens resultantes do processo de segmentação do ortomosaico; as imagens são anotadas, caracterizando-se como aprendizagem supervisionada.

Na sexta etapa os classificadores são testados com 40% das imagens não submetidas ao processo de anotação; os resultados finais de contagem de animais da área são apresentados e contrastados com os registros de presença de animais na área de interesse para avaliação dos resultados da classificação, em relação a métrica acurácia.

As etapas acima foram integralmente aplicadas em dois experimentos e parcialmente no terceiro, conforme detalhado a seguir.

### 5.2.1 Primeiro experimento

1. Na primeira etapa foi escolhida uma das imagens (com animais) em seu formato original, divididas em imagens de 400 por 400 pixels (px), e a partir destas foram selecionados 52 animais - definidos como classe vaca - e anotados com sua respectiva posição na imagem (x e y) - chamada *bounding box* (Figura 17). Como técnica para aumento dos dados de treinamento, estas imagens foram rotacionadas a 45, 90, 135 e 180 graus.

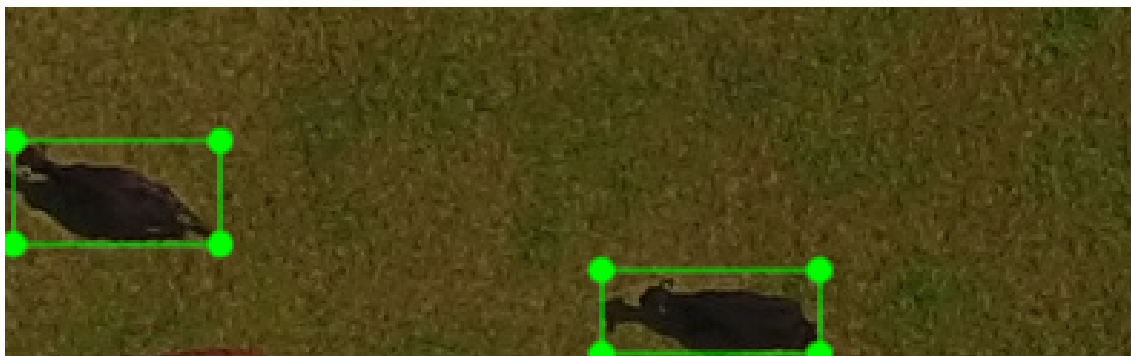
Estas anotações de animais foram definidas como conjunto de treinamento para a rede neural convolucional, representando 60% dos 87 animais encontrados no aerolevanteamento. O conjunto de imagens de teste foi dos 40% restantes (todos os experimentos deste trabalho usaram a mesma proporção).

2. A imagem total da área foi dividida em tamanhos de 200 px até 800 px quadrados, com intervalos de 100 px, com a intenção de diminuir o tamanho total da imagem e identificar as variações na acurácia.
3. Dentro do Colab foi criado um script para instalação do TensorFlow 1.14, importação de imagens de treinamento e classificação.
4. As imagens da etapa dois foram submetidas ao algoritmo classificador *Faster Region Convolutional Neural Network (Faster RCNN)*.

### 5.2.2 Segundo experimento

Para o segundo experimento, foi realizado outro aerolevanteamento, as 12 h (condição teoricamente ideal de posição solar), com 50 animais dispostos de forma aleatória na

Figura 17 – Imagem com anotações em animais



Fonte: Autor (2019)

área, sendo que estes são outros animais em relação ao primeiro experimento, mas desta vez em posições em pé e deitado, com as mesmas configurações da CNN do primeiro experimento, mas criando uma classe complementar denominada “vaca deitada”. Dos 60% (30 animais) das imagens necessárias para treinamento, o resultado gerado no primeiro experimento foi capaz de automatizar a classificação de 44%. Os 16% restantes foram criados por meio de anotação manual das *bounding boxes*.

### 5.2.3 Terceiro experimento

As imagens do primeiro e do segundo experimento foram submetidas em conjunto ao treinamento da rede neural, mantendo as mesmas configurações e proporções, com o objetivo de aumentar a acurácia do primeiro experimento e avaliar se há influência da quantidade de imagens.

### 5.2.4 Quarto experimento

Como este método foi proposto como uma alternativa não invasiva aos animais, realizou-se um experimento para avaliar se mesmo a baixas altitudes (10 m) de voo sobre os animais haveria algum tipo de reação por parte deles com a proximidade da aeronave. Ou seja, avaliar se, com a aplicação do método, próximo aos animais, haveria interferência perceptível no cotidiano destes.

Do ponto de partida da área experimental foi feito um voo manual até 50 m de altitude, e então foram efetuadas cinco tentativas de sobrevoo até 10 m registrando as

imagens do conjunto de nove animais disponíveis.

### 5.2.5 Desempenho computacional dos experimentos

Para verificar se este método teria condições de ser executado, de forma remota, na etapa de treinamento dos classificadores, a métrica tempo de execução (tempo decorrido entre o início do treinamento até o fim em cada ambiente - CPU e GPU) foi avaliada em relação ao número de épocas e variação da acurácia. Mais precisamente, neste experimento o método foi submetido à uma variação na quantidade de épocas (de 600 a 1200) no ambiente Google Colab, alternando entre CPU e GPU, observando-se o efeito na acurácia e no tempo de execução.

### 5.3 Métricas

Para avaliar a eficácia do método, adotou-se as duas formas de avaliação anteriormente definidas: acurácia e *Intersection over Union* (IoU). Complementarmente, foi avaliado o tempo de execução como métrica de desempenho associada às variantes do ambiente computacional utilizado nos experimentos.

A métrica **acurácia** é calculada conforme a Equação 2:

$$\frac{VPR + VN}{VP + FP + VN + FN} \quad (2)$$

Onde, Verdadeiro Positivo Real (VPR): foi detectado corretamente excluindo dupla detecção (DD); Verdadeiro Negativo (VN): não foi detectado corretamente; Verdadeiro Positivo (VP): foi detectado corretamente; Falso Positivo (FP): foi detectado, mas não é um animal; Falso Negativo (FN): existe e não foi detectado corretamente.

Já a métrica **IoU** mede a taxa de sobreposição, isto é, quanto há de sobreposição entre a região predita e a região verdadeira do objeto a ser buscado. A Equação 3 define área de sobreposição (AS) e área de união (AU):

$$\frac{AS}{AU} \quad (3)$$

A última métrica avaliada foi **tempo de execução**, mensurada em minutos, relacionada ao tempo decorrido entre o início e o final da etapa do método, com a variação



de épocas e acurácia nas unidades de processamento CPU e GPU disponíveis no Google Colab.

## 6 RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados e discussões acerca dos experimentos detalhados na Seção 5.2. Os resultados são apresentados na forma de tabelas, gráficos e ilustrações. As três tabelas possuem oito colunas, representando os seguintes dados:

- Total: total de animais presentes na área durante o aerolevamento passíveis de serem identificados;
- Verdadeiro Positivo Real (VPR): foi detectado corretamente;
- Verdadeiro Negativo (VN): não foi detectado corretamente;
- Falso Positivo (FP): foi detectado, mas não é um animal;
- Falso Negativo (FN): existe e não foi detectado corretamente;
- Dupla Detecção (DD): animal detectado com duplicidade;
- *Intersection over Union* (IoU): taxa de sobreposição;
- Acurácia: percentual de animais detectados corretamente.

O resultado do Experimento 1, conforme ilustrado na Tabela 1, mostra que o método foi capaz de detectar com uma acurácia de 86% com 1000 épocas e sobreposição (IoU) mínima de 75% de intersecção correta entre as *bounding boxes*. A variação de IoU interfere na acurácia do método, já que o animal pode estar fragmentado entre duas imagens separadas. Outra tentativa foi feita para aumentar a acurácia diminuindo o IoU para 65%, e por fim a 60%, sendo obtida a acurácia de 91%.

Tabela 1 – Experimento 1 incluindo imagens do primeiro voo com 1000 épocas

Total	VP	FP	VN	FN	DD	IoU	Acurácia
35	30	0	0	5	0	75%	86%
35	31	0	0	4	0	65%	88%
35	29	0	0	5	0	60%	91%

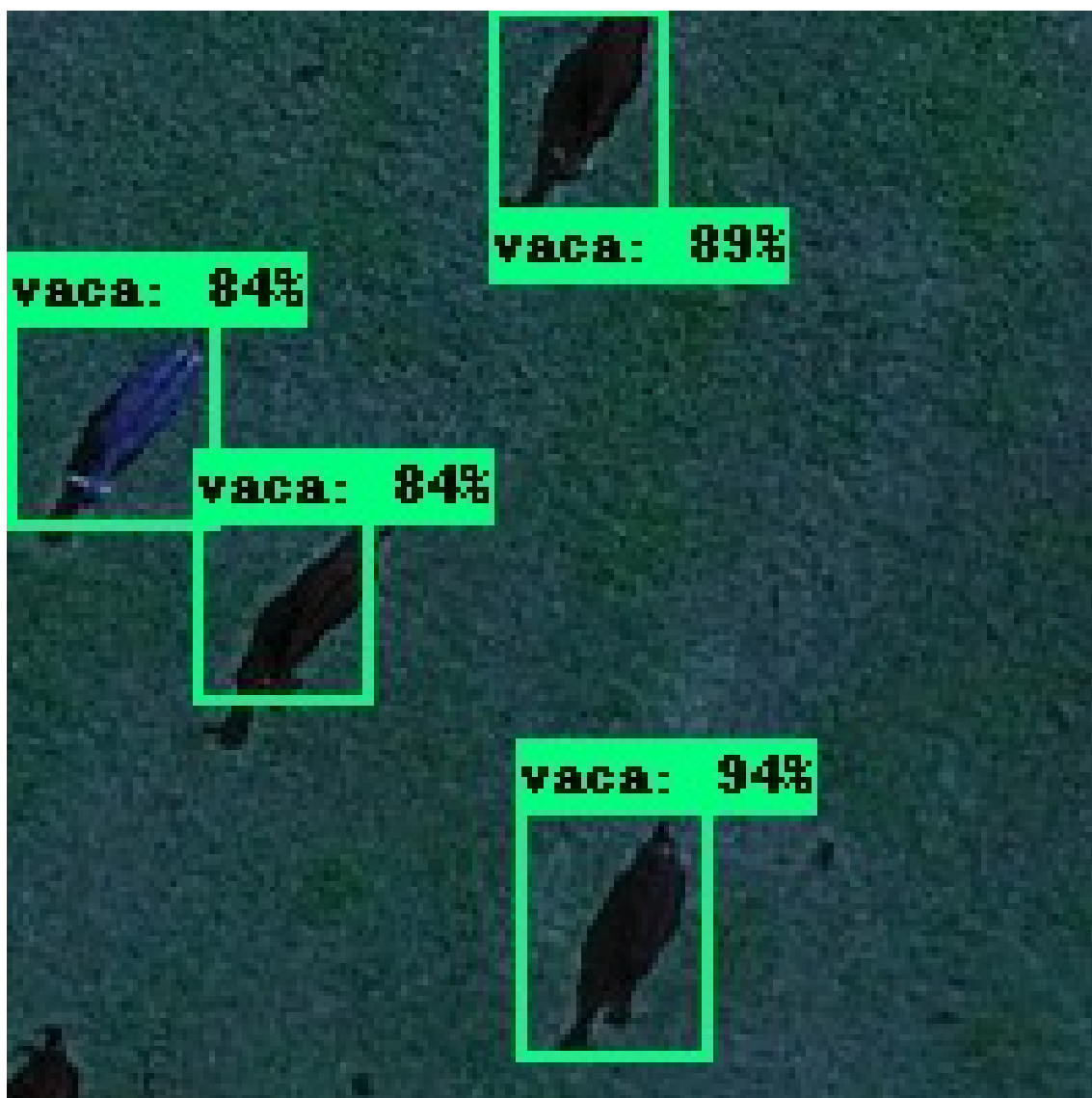
Fonte: Autor (2019)

Na Figura 18 pode ser observada a detecção da classe vaca e variação de IoU em cada animal.

Já no Experimento 2, conforme os resultados apresentados na Tabela 2, o método foi capaz de detectar animais com uma acurácia de 92%, com 1000 épocas e IoU de 60%. Com esse segundo experimento, foi possível confirmar a tendência de impacto do IoU, isto é, que o funcionamento com método com IoU menor melhora a acurácia.

Na Figura 19 pode ser observada a detecção da classe vaca deitada e variação de

Figura 18 – Imagem com detecção e localização de animais referente ao Experimento 1



Fonte: Autor (2019)

Tabela 2 – Experimento 2 incluindo imagens do segundo voo com 1000 épocas

Total	VPR	FP	VN	FN	DD	IoU	Acurácia
22	17	0	0	5	0	75%	77%
22	19	0	0	3	0	65%	86%
22	20	1	0	1	0	60%	92%

Fonte: Autor (2019)

IoU em cada animal. Em particular, é possível perceber um caso no qual parte da cerca foi detectada como vaca (FP). Um aspecto relevante desse experimento é a hora do voo (12 h), não havendo interferência significativa de sombras.

O Experimento 3, mostrado na Tabela 3, teve a acurácia de 92%, mas houve au-

Figura 19 – Imagem com detecção e localização de animais referente ao Experimento 2



Fonte: Autor (2019)

mento na contagem de animais com a variação do IoU. Neste experimento ocorreu uma contagem duplicada. Quando ocorreram duplicações de detecção de animais, foi necessário incluir um critério de exclusão para os casos nos quais há sobreposição de pelo menos 50% entre as *bounding boxes* no mesmo animal.

Tabela 3 – Experimento 3 incluindo imagens dos dois voos com 1000 épocas

Total	VPR	FP	VN	FN	DD	IoU	Acurácia
28	25	0	0	3	0	75%	90%
28	26	0	0	2	0	65%	92%
28	26	0	0	2	0	60%	92%

Fonte: Autor (2019)

Ao executar o Experimento 4, notou-se que não houve qualquer movimentação significativa (raio de 1 m) dos animais com o sobrevoo do VANT a 10 m de altitude, inclusive quando estiveram dispostos unidos, como é visualizado na Figura 20.

Figura 20 – Imagem com 18 animais a 10 m de altitude referente ao Experimento 4



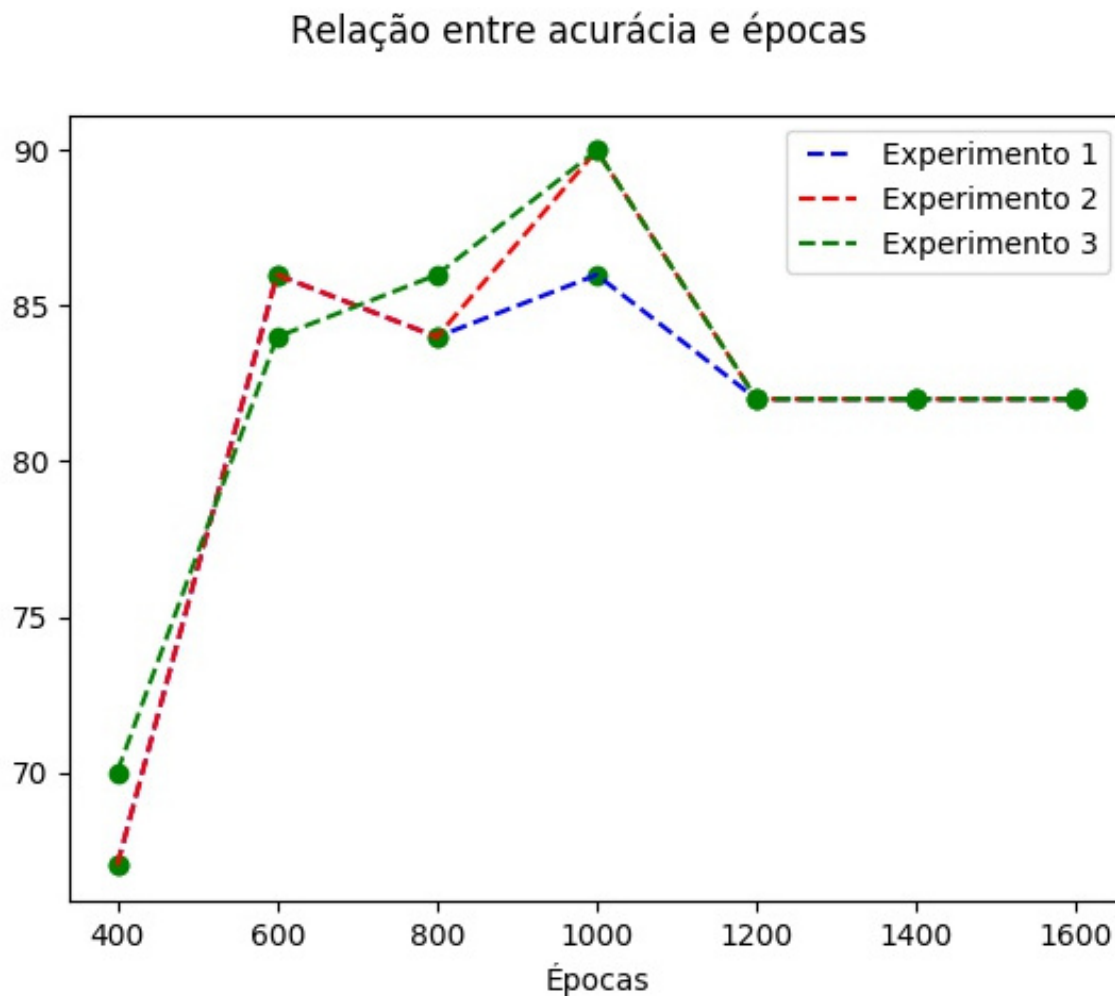
Fonte: Autor (2019)

A Figura 21 apresenta os resultados do experimento complementar, necessário para identificar a quantidade adequada de épocas para término do treinamento, considerando os cenários dos três experimentos de classificação. É possível concluir que a melhor acurácia foi obtida com 1000 épocas, sendo que a partir de 1200 épocas houve queda na capacidade de classificação adequada dos animais.

Todos os experimentos apresentaram maior acurácia em 300px e 400px em razão do tamanho das imagens e capacidade de detecção da rede neural. A Figura 22 ilustra a variação nos Experimentos 1, 2 e 3, sendo possível observar que, quanto menor a IoU do objeto (animal) a ser encontrado, maior tende a ser a eficácia da contagem. O IoU é fator determinante para acurácia já que é possível encontrar animais abaixo de 100% de sobreposição, levando a um resultado eficaz com animais vistos sob outros ângulos, característica que ocorre em aerolevantamentos.

A partir das *bounding boxes* obtidas no primeiro voo, foi possível identificar a localização geográfica aproximada dos animais sobrepondo o arquivo JPG sobre o arquivo TIFF (arquivo que contém informações sobre as posições geográficas) correspondente no ortomosaico. A Figura 23 foi criada a partir de um mapa temático retirado do Google

Figura 21 – Relação entre acurácia e épocas

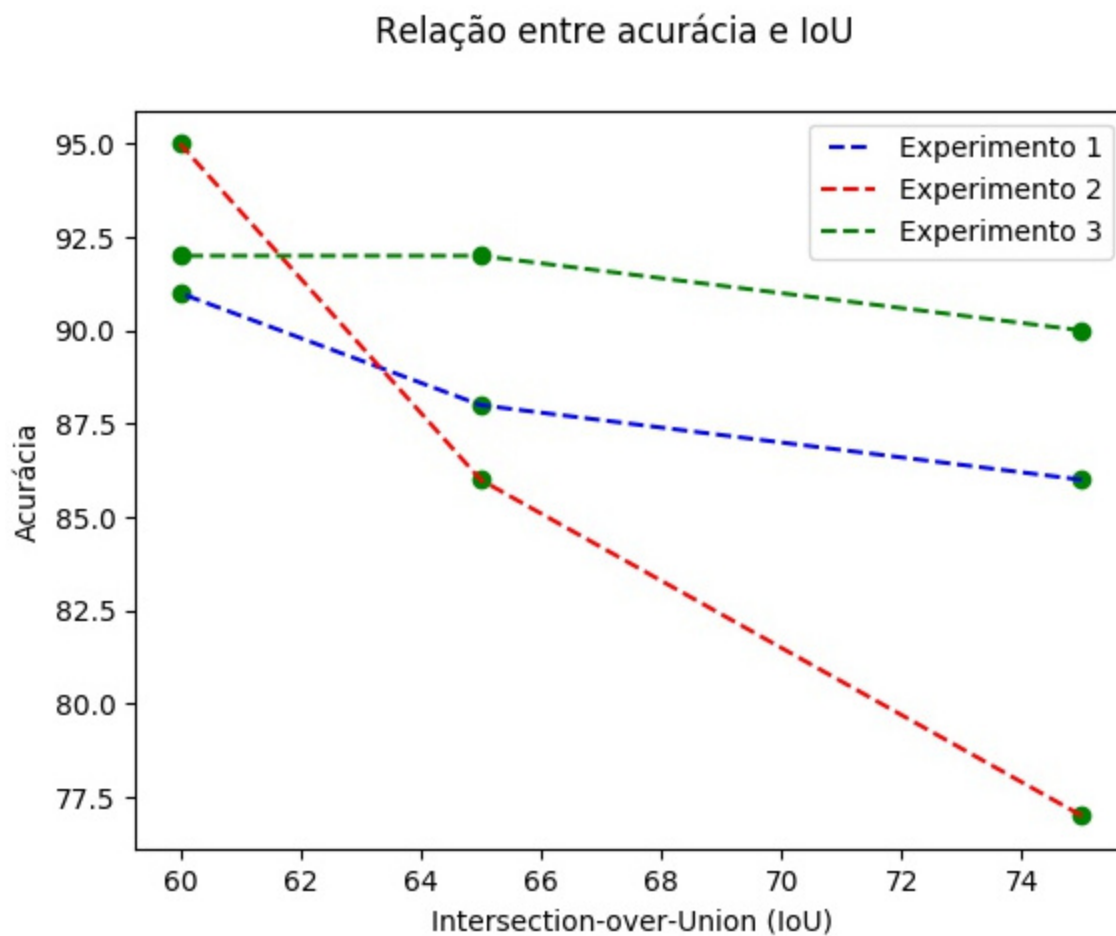


Fonte: Autor (2019)

Maps, correspondente à área na qual foi realizado o voo do Experimento 1. As posições geográficas foram retiradas das imagens e identificadas pelos ícones azuis, facilitando a compreensão da distribuição dos animais.

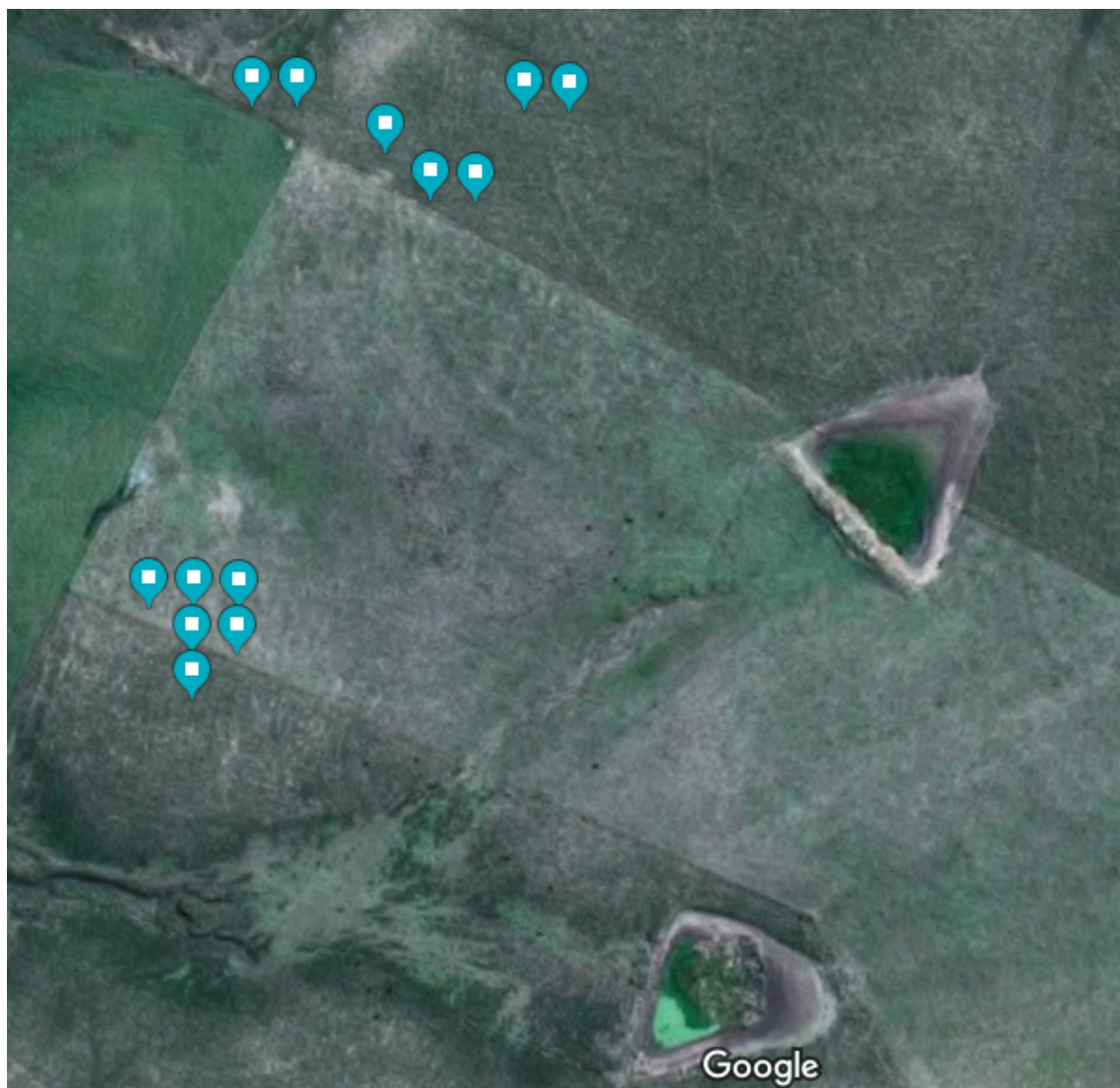
No que se refere ao desempenho computacional do método, conforme pode ser visto na Figura 24, o eixo Y destaca a métrica tempo no treinamento de classificação no ambiente 1 (GPU) e ambiente 2 (CPU). Aplicando-se a variância populacional entre os tempos, tanto do ambiente 1 quanto 2, chega-se a variância de 850,74, indicando uma discrepância grande entre os valores. Considerando a necessidade de aumentar a acurácia para execução do método, conclui-se que o uso de uma GPU é o mais indicado.

Figura 22 – Relação entre a melhora da acurácia e IoU



Fonte: Autor (2019)

Figura 23 – Imagem com localização geográfica aproximada dos animais

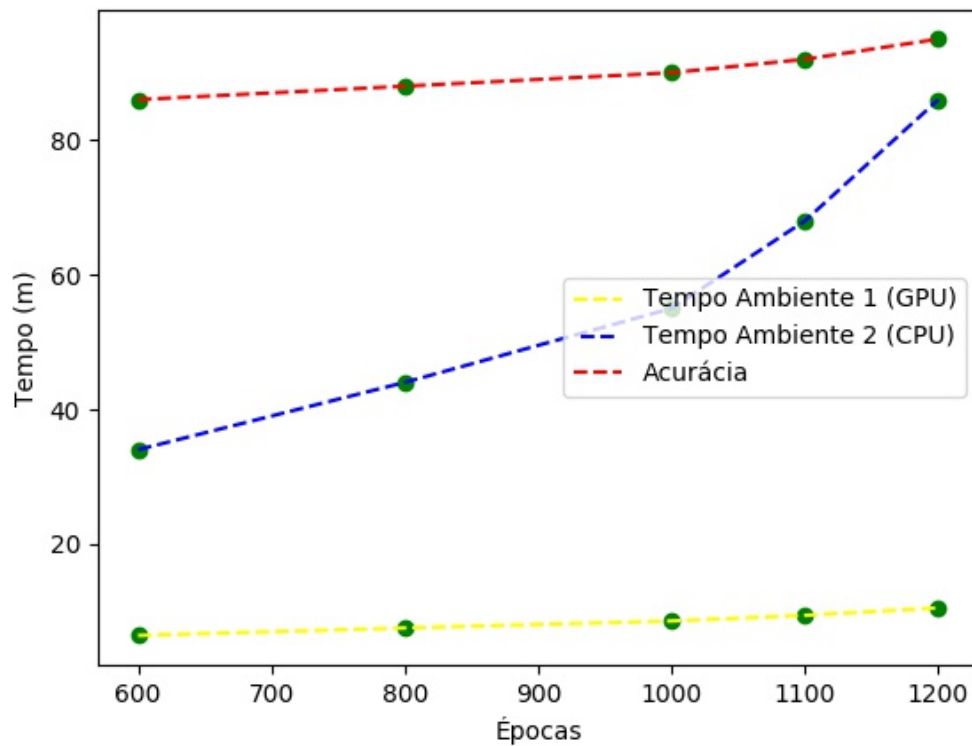


Fonte: Autor (2019)



Figura 24 – Relação entre acurácia e tempo nos ambientes 1 (GPU) e 2 (CPU)

Relação entre tempo, acurácia e épocas ambiente 1 (GPU) e 2 (CPU)



Fonte: Autor (2019)

## 7 CONCLUSÕES

Este trabalho, por meio da aplicação da visão computacional em pecuária de manejo extensivo, alcançou a contagem de animais automatizada por meio de aerolevanteamento via VANT com até 92% de acurácia. Além disso, o método proposto foi capaz de indicar a distribuição espacial aproximada dos animais. Portanto, mostrou ser possível minimizar a interferência humana no processo de contagem de animais em áreas de dezenas de hectares, como as existentes nos campos experimentais da Embrapa Pecuária Sul.

No contexto de visão computacional, a extração de características como bordas, retirada de ruídos e rotação dos objetos por meio das CNN, permitiu diminuir as atividades de pré-processamento e aumentar a acurácia da localização e detecção dos animais, mesmo sem um treinamento inicial da CNN. Com o método apresentado (não invasivo) foi possível identificar animais à distância de 100 m e contá-los a partir de imagens RGB registradas por VANT. O método demonstrou melhores resultados quando animais estiveram em posição deitada e com maior incidência de luz solar. Embora a contagem de animais não tenha se aproximado de 100%, a contagem foi próxima ao total, demonstrando que o método é capaz de ser usado na estimativa de animais, mas pode ser aprimorado coletando mais imagens de outros voos em séries temporais, inclusive em outras áreas. Quando avaliada a métrica tempo de execução do treinamento da CNN com as imagens, e o aumento das épocas para melhorar a acurácia, tornou-se importante o uso de uma GPU, já que reduzir tempo é necessário para a automatização da execução do método.

Este método contempla outras aplicações como encontrar animais perdidos em áreas vizinhas da propriedade que está sendo mapeada. Obter uma estimativa em áreas com conjunto acima de 50 animais, o que não é viável fazer com essa quantidade de animais de forma manual. Outra forma de contagem manual é a contenção de animais em uma determinada área. No entanto, deslocar animais à campo pode trazer perda de peso destes, trazendo prejuízo ao produtor rural.

Além destas reflexões e considerações mais amplas sobre o trabalho desenvolvido, mostra-se necessário concluir sobre o processo investigativo realizado, a luz das quatro principais questões de pesquisa delineadas:

- A qualidade das imagens RGB coletadas por meio de um VANT é suficiente para identificar os animais a campo? Sim, o método proposto e os resultados obtidos com a metodologia experimental adotada apontaram para a viabilidade deste tipo de

câmera, para cenários com incidência de luz solar.

- É possível mensurar o impacto de fatores que influenciam a coleta de imagens por meio de VANT, tais como a altitude (maior ou menor resolução espacial), o terreno (plano, ondulado ou acidentado), a vegetação (presença de árvores e arbustos) e a luminosidade (horário e nebulosidade), na eficácia do método de identificação dos animais? Parcialmente. A metodologia experimental adotada (restrita a uma única altitude, apenas com terreno suavemente ondulado e vegetação com mínima presença de árvores e arbustos) não foi suficiente para identificar a influência da maior parte destes fatores. Contudo, envolveu voos em dois horários, permitindo concluir que voos no horário do meio dia reduzem a ocorrência de sombras, o que aumentou a acurácia do método.
- A visão computacional consegue suprir a necessidade da contagem de animais sob a perspectiva de processamento computacional? Sim, as técnicas utilizadas, com a aplicação de recursos computacionais comuns nos dias atuais - incluindo processadores *multicore* e *graphical processing units* (GPU) -, foram suficientes para executar o método em tempo hábil para diferentes tipos de aplicações, em especial para a contagem do rebanho em uma área de dezenas de hectares.
- Há técnicas capazes de realizar essa tarefa com acurácia maior que 90%, dentro do contexto de visão computacional? Sim, os resultados experimentais mostraram diversos casos nos quais foi obtida acurácia superior a este limite.

Entre as limitações observadas no método, destaca-se a relacionada ao uso de câmera RGB capaz de reconhecer apenas objetos no espectro visível, o que o torna suscetível a falhas, por exemplo, quando animais estão abaixo de árvores ou arbustos. Como possível solução de contorno para este problema, seria relevante realizar voos ao nascer do sol, período no qual os animais não buscam regiões de sombra para se proteger. Outro aspecto importante é a relação entre a variação da altitude e tamanho na imagem. Como forma de resolver esta questão, faz-se necessário coletar imagens a diversas altitudes em relação ao nível do mar, buscando manter constante a distância entre o elemento imageador (câmera) e o objeto (animal), adaptando-se à ondulação do terreno. Um fator extra é o processo de criação do ortomosaico, introduzido no método com o objetivo de diminuir a contagem duplicada de animais. No entanto, constatou-se que houve movimentação de animais, implicando no registro de imagens fantasma, carentes de um tratamento complementar para otimizar a capacidade de evitar a dupla contagem.

Como trabalhos futuros pretende-se registrar imagens através de vídeo com câmera RGB, aliando técnicas de reconhecimento de objetos (CNN) a técnicas de *tracking*, mantendo a relação entre espaço e tempo ao longo do voo. Além deste, cogita-se o uso de câmeras térmicas baseadas no princípio de que os animais emitem ondas com variação de temperatura corporal, sendo translúcidos a obstáculos. Também mostra-se pertinente buscar alternativas capazes de estimar o peso vivo dos animais como complemento aos métodos convencionais de controle de estoque de propriedades com manejo extensivo. Considerando que este método envolve significativo volume de dados e processamento intensivo, pretende-se avaliar, por meio de métricas de desempenho e de comunicação, a utilização de diferentes arquiteturas computacionais para o armazenamento e processamento das imagens, tendo como extremos uma arquitetura convencional centralizada e uma arquitetura avançada para Computação em Névoa (*Fog Computing*).

Por fim, cabe destacar a percepção sobre o potencial do método proposto para estudos que visem uma melhor compreensão acerca do comportamento do rebanho. Potencial este que carece de uma metodologia experimental própria, com delineamento de hipóteses e experimentos específicos para comprová-las, caracteriando-se como um relevante trabalho futuro de pesquisa aplicada.

## REFERÊNCIAS

- AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **Regulamentação da Agência Nacional de Aviação**. Brasília, DF, 2017. Disponível em: <<http://www.anac.gov.br/assuntos/legislacao/legislacao-1/rbha-e-rbac/rbac/rbac-e-94-emd-00>>. Acesso em: 10 jan. 2019.
- ALVES, M. A. Z.; FREITAS, H.; NAVAUX, P. Ensino de arquiteturas de processadores many-core e memórias cache utilizando o simulador simics. In: \_\_\_\_\_. [S.l.: s.n.], 2012. p. 74–110. ISBN 9788576692638.
- ANDREW, W.; GREATWOOD, C.; BURGHARDT, T. Visual localisation and individual identification of holstein friesian cattle via deep learning. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2017. p. 2850–2859.
- ARAÚJO, T. M. U. de *et al.* Treinamento de uma rede perceptron de múltiplas camadas utilizando enhanced-continuous-grasp. 2015.
- AZQUETA-ALZÚAZ, A. *et al.* Massive data load on distributed database systems over hbase. In: IEEE. **Cluster, Cloud and Grid Computing (CCGRID), 2017 17th IEEE/ACM International Symposium on**. [S.l.], 2017. p. 776–779.
- BARBALHO, V. F. *et al.* O controle de estoque de animais na pecuária bovina de corte: uma questão de continuidade. In: **Anais do Congresso Brasileiro de Custos-ABC**. [S.l.: s.n.], 2005.
- BEZERRA, E. **Introdução à Aprendizagem Profunda**. [S.l.: s.n.], 2016.
- CARNEIRO, T. *et al.* Performance analysis of google colab as a tool for accelerating deep learning applications. **IEEE Access**, IEEE, v. 6, p. 61677–61685, 2018.
- CHAMOSO, P. *et al.* Uavs applied to the counting and monitoring of animals. In: **Ambient Intelligence-Software and Applications**. [S.l.]: Springer, 2014. p. 71–80.
- DENG, J. *et al.* Imagenet: A large-scale hierarchical image database. In: **In CVPR**. [S.l.: s.n.], 2009.
- DERRAS, B.; BEKKOUCHE, A.; ZENDAGUI, D. Neuronal approach and the use of kik-net network to generate response spectrum on the surface. **Jordan Journal of Civil Engineering**, v. 4, p. 12–21, 2010.
- DU, N. *et al.* Spatio-temporal data index model of moving objects on fixed networks using hbase. In: IEEE. **Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on**. [S.l.], 2015. p. 247–251.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **EMBRAPA**. Porto Alegre, RS, 2018. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/26426446/pecuaria-de-precisao-ganha-destaque-na-40-expo>>. Acesso em: 10 jan. 2018.

FERREIRA, A. dos S. **Redes Neurais Convolucionais Profundas na Detecção de Plantas Daninhas em Lavoura de Soja**. 70 p. Dissertação (Masters Thesis) — Universidade Federal do Mato Grosso do Sul, Campo Grande, March 2017.

FERREIRA, J. R. J.; OLIVEIRA, M. C. Banco de dados nosql público de nódulos pulmonares para auxílio à pesquisa e diagnóstico do câncer de pulmão. In: **XXIV Congresso Brasileiro de Engenharia Biomédica (CBEB 2014)**. [S.l.: s.n.], 2014. p. 177–180.

FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA. **Características da Agropecuária do RS**. Porto Alegre, RS, 2017. Disponível em: <<https://www.fee.rs.gov.br/sinteseilustrada/%20caracteristicas-da-agropecuaria-do-rs>>. Visitado em nov., 2017.

GEMERT, J. C. van *et al.* Nature conservation drones for automatic localization and counting of animals. In: SPRINGER. **Workshop at the European Conference on Computer Vision**. [S.l.], 2014. p. 255–270.

GIRSHICK, R. Fast r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2015. p. 1440–1448.

HANDCOCK, R. N. *et al.* Monitoring animal behaviour and environmental interactions using wireless sensor networks, gps collars and satellite remote sensing. **Sensors**, Molecular Diversity Preservation International, v. 9, n. 5, p. 3586–3603, 2009.

HU, R.; COLLOMOSSE, J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. **Computer Vision and Image Understanding**, Elsevier, v. 117, n. 7, p. 790–806, 2013.

HU, W. *et al.* Deep convolutional neural networks for hyperspectral image classification. **Journal of Sensors**, Hindawi, v. 2015, 2015.

JENSEN, J. R.; EPIPHANIO, J. C. N. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. [S.l.]: Parêntese, 2011.

LIU, W. *et al.* Ssd: Single shot multibox detector. In: SPRINGER. **European conference on computer vision**. [S.l.], 2016. p. 21–37.

LIU, Y. *et al.* Massive image data management using hbase and mapreduce. In: IEEE. **Geoinformatics (GEOINFORMATICS), 2013 21st International Conference on**. [S.l.], 2013. p. 1–5.

LOHAR, N. *et al.* Content based image retrieval system over hadoop using mapreduce. 2016.

MENESES, P. R.; ALMEIDA, T. d. e. o. Introdução ao processamento de imagens de sensoriamento remoto. **Brasília: UNB/CNPq**, 2012.

MIAO, X. *et al.* Insulator detection in aerial images for transmission line inspection using single shot multibox detector. **IEEE Access**, IEEE, 2019.

MOROZ, R. C.; JR, S. L. S. Estado da arte de sensores aplicados a pecuária bovina. In: CONFEA. **Congresso Técnico Científico da Engenharia e da Agronomia – CONTECC'2016**. [S.l.], 2016.

MULERO-PÁZMÁNY, M. *et al.* Unmanned aircraft systems complement biologging in spatial ecology studies. **Ecology and evolution**, Wiley Online Library, v. 5, n. 21, p. 4808–4818, 2015.

NEPOMUCENO, P. L. M. **Geoprocessamento nos mapeamentos de inventário dos Planos de Manejo de Unidades de Conservação paulistas: avaliação dos casos do PETAR e da APA-VRT a partir da revisão integrada da literatura**. São Paulo, 2016. 122 p.

NGUYEN, L. D. *et al.* Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation. In: IEEE. **Circuits and Systems (ISCAS), 2018 IEEE International Symposium on**. [S.l.], 2018. p. 1–5.

PHAM, B. T. *et al.* Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at himalayan area (india) using gis. **Catena**, Elsevier, v. 149, p. 52–63, 2017.

POLIDORIO, A. M. *et al.* Detecção automática de sombras e nuvens em imagens cbers e landsat 7 etm. **Simpósio Brasileiro de Sensoriamento Remoto, Brasil**, p. 4233–4240, 2005.

POLONI, K. M. Identificação de direção de imageamento em vídeos aéreos. 2016.

PRADEEP, J.; SRINIVASAN, E.; HIMAVATHI, S. Diagonal based feature extraction for handwritten character recognition system using neural network. In: IEEE. **Electronics Computer Technology (ICECT), 2011 3rd International Conference on**. [S.l.], 2011. v. 4, p. 364–368.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2ª Edição**. [S.l.]: Editora Feevale, 2013.

RAGOTHAMAN, S. *et al.* Unsupervised segmentation of cervical cell images using gaussian mixture model. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops**. [S.l.: s.n.], 2016. p. 70–75.

REY, N. Combining uav-imagery and machine learning for wildlife conservation. 2016.

RUSSAKOVSKY, O. *et al.* Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, Springer, v. 115, n. 3, p. 211–252, 2015.

SAMUELSSON, O. **Video Tracking Algorithm for UnmannedAerial Vehicle Surveillance**. 2012.

SEDAGHAT, A.; EBADI, H. Remote sensing image matching based on adaptive binning sift descriptor. **IEEE transactions on geoscience and remote sensing**, IEEE, v. 53, n. 10, p. 5283–5293, 2015.

SILVA, C. *et al.* Utilização de vant para geração de ortomosaicos e aplicação do padrão de exatidão cartográfica (pec). **Artigo apresentado no SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO (SBSR)**, 2015.

SILVA, D. C. D. *et al.* Qualidade de ortomosaicos de imagens de vant processados com os softwares aps, pix4d e photoscan. **V Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, Recife-PE**, 2014.

SILVA, F.; SALVIANO, A.; ANDRADE, J. d. B. Áreas degradadas em microbacia de gilbués-pi utilizando imagens do sensor ccd-cbers-2. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 13, p. 4257–4260, 2007.

SILVA, M. A. Oliveira da; LIPORACE, F. dos S. Detecção automática de nuvem e sombra de nuvem em imagens de sensoriamento remoto. **Boletim de Ciências Geodésicas**, Universidade Federal do Paraná, v. 22, n. 2, 2016.

TEIXEIRA, J. C.; HESPANHOL, A. N. A trajetória da pecuária bovina brasileira. **Caderno Prudentino de Geografia**, v. 1, n. 36, p. 26–38, 2014.

WANG, S. *et al.* Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic abc and biogeography-based optimization. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 17, n. 8, p. 5711–5728, 2015.

WOLF, P. R. **Solutions Manual to Accompany Elements of Photogrammetry: With air photo interpretation and remote sensing**. [S.l.]: McGraw-Hill, 1983.

XIAO, Z.; LIU, Y. Remote sensing image database based on nosql database. In: IEEE. **Geoinformatics, 2011 19th International Conference on**. [S.l.], 2011. p. 1–5.

YANG, K.; GENG, F. Application of faster r-cnn model on human running pattern recognition. **arXiv preprint arXiv:1811.05147**, 2018.

ZHAO, Y. *et al.* Low-altitude uav imagery based cross-section geological feature recognition via deep transfer learning. In: IEEE. **2018 3rd International Conference on Robotics and Automation Engineering (ICRAE)**. [S.l.], 2018. p. 253–257.